

1956 - 2006

TRƯỜNG ĐẠI HỌC BÁCH KHOA HÀ NỘI
50 NĂM XÂY DỰNG VÀ PHÁT TRIỂN

KHOA TOÁN - TIN ỨNG DỤNG
PGS. TS. BÙI MINH TRÍ

Kinh tế lượng



NHÀ XUẤT BẢN KHOA HỌC VÀ KỸ THUẬT



1956 - 2006

TRƯỜNG ĐẠI HỌC BÁCH KHOA HÀ NỘI

50 NĂM XÂY DỰNG VÀ PHÁT TRIỂN

KHOA TOÁN – TIN ỨNG DỤNG

PGS. TS. BÙI MINH TRÍ

KINH TẾ LƯỢNG



NHÀ XUẤT BẢN KHOA HỌC VÀ KỸ THUẬT

HÀ NỘI

LỜI NÓI ĐẦU

Kinh tế lượng là một môn khoa học trong đó áp dụng các công cụ lý thuyết kinh tế toán học và suy đoán thống kê để phân tích các vấn đề kinh tế. Nếu như kinh tế vĩ mô mô tả sự vận động của toàn bộ nền kinh tế, kinh tế vi mô mô tả hành vi của người sản xuất và tiêu dùng thì kinh tế lượng trang bị cho nhà kinh tế một phương pháp lượng hoá và phân tích sự vận động và các hành vi trên để đưa ra mô hình phù hợp nhất cho một hiện tượng kinh tế nào đó.

Chính vì vậy từ khi ra đời cho đến nay, kinh tế lượng đã đem lại cho các nhà kinh tế một công cụ đo lường sắc bén để đo lường các quan hệ kinh tế. Ngày nay phạm vi sử dụng của kinh tế lượng đã vượt quá phạm vi kinh tế, đã lan sang các lĩnh vực khác như xã hội học, vũ trụ học...

Sự đòi hỏi của kinh tế lượng trong phân tích định lượng các hiện tượng kinh tế, kiểm định sự phù hợp và độ tin cậy của các giả thiết trong quá trình hoạch định các chính sách vĩ mô cũng như ra các quyết định tác nghiệp, việc dự báo và dự đoán có độ tin cậy cao..., tất cả những điều đó đã làm cho kinh tế lượng có một vai trò ngày càng quan trọng và bản thân nó cũng không ngừng được phát triển và hoàn thiện.

Sự phát triển của máy tính điện tử trong khoa học nói chung và kinh tế lượng nói riêng đã làm tăng sức mạnh của kinh tế lượng. Chính điều này giúp các nhà kinh tế kiểm chứng được các lý thuyết kinh tế có thích hợp hay không, dẫn tới những quyết định đúng đắn đem lại hiệu quả tối ưu cho các hoạt động kinh doanh, hoạch định được các chính sách và chiến lược tốt trong kinh tế xã hội.

Khi nghiên cứu cơ sở Kinh tế lượng ta sẽ thực sự thấy say mê và hứng thú. Việc tìm hiểu phương pháp luận của Kinh tế lượng giúp chúng ta tìm được mô hình tổng quát của một hoạt động kinh tế dựa trên những bộ số liệu đầu vào, tìm ra mối quan hệ giữa các số liệu đầu vào, phân tích đánh giá để tìm ra mô hình đúng đắn và sát thực cho hoạt động kinh tế. Thông qua đó, ta sẽ biết quá trình vận động, quá trình tương tác lẫn nhau giữa các hiện tượng kinh tế, cũng như các hiện tượng khoa học tự nhiên và khoa học xã hội.

Sự đòi hỏi của Kinh tế lượng trong phân tích định lượng các các hiện tượng kinh tế, kiểm định sự phù hợp và độ tin cậy của các giả thiết trong quá trình hoạch định các chính sách vĩ mô cũng như ra các quyết định tác nghiệp, việc dự báo và dự đoán có độ tin cậy cao,... , tất cả những điều đó đã làm cho Kinh tế lượng có một vai trò ngày càng quan trọng và bản thân nó cũng không ngừng được phát triển và hoàn thiện.

Nội dung Giáo trình bao gồm 6 chương:

Chương 1. Những khái niệm cơ bản và phương pháp luận của kinh tế lượng

Nêu ra những khái niệm và phương pháp luận của kinh tế lượng, đồng thời cũng nhắc lại những kiến thức về xác suất thống kê cần dùng để tính toán và kiểm định.

Chương 2. Mô hình hồi quy hai biến, ước lượng và kiểm định giả thuyết

Trong chương này, trình bày thế nào là phân tích hồi quy, nêu ra những loại số liệu thu thập được để phân tích hồi quy và sau đó đưa ra được mô hình hồi quy. Nội dung chính của chương này bước đầu làm quen với mô hình hồi quy hai biến, tạo cơ sở và tiền đề cho việc nghiên cứu mô hình hồi quy tổng quát.

Chương 3. Hồi quy bội

Đây là chương đưa ra mô hình hồi quy tổng quát, dùng phương pháp bình phương cực tiểu (BPCT) để ước lượng các hệ số của mô hình, sau đó kiểm định tính xem mô hình đưa ra có phù hợp hay không?

Chương 4. Một số trường hợp mở rộng của hồi quy bội

Để ra được mô hình cho một hình thức kinh tế , người ta đã đưa ra một số giả thiết cho việc xây dựng mô hình đó. Trong chương này sẽ lần lượt nêu ra mô hình không thoả mãn một trong các giả thiết đó, cách phát hiện ra các giả thiết không phù hợp đó. Trong chương này, cũng đưa ra một số sai lầm và phát hiện một số sai lầm mắc phải khi lập mô hình.

Chương 5. Quy hoạch trực giao

Trong chương này, đưa ra những khái niệm, tính chất và cách xây dựng mô hình trực giao. Ở đây xây dựng cả mô hình trực giao cấp một và cấp hai.

Quy hoạch trực giao (QHTG) dựa trên phương pháp BPCT- chỉ khác là chủ động bố trí thí nghiệm để thu được ma trận đầu vào như ý muốn. Giúp xây dựng được và tính toán được với các mô hình nhiều biến.

Chương 6. Áp dụng mô hình hồi quy vào phương pháp ngoại suy để dự báo

Chương này giới thiệu qua về dự báo, đồng thời nêu ra phương pháp ngoại suy để dự báo.

Phụ lục

Phụ lục 1: Đưa vào các bảng phân vị cần dùng.

Phụ lục 2: Chương trình giải bài toán Kinh tế lượng.

Phụ lục 2: Một số kết quả tính toán trên máy tính.

Tác giả xin chân thành cảm ơn độc giả góp ý kiến để nâng cao chất lượng cuốn sách.

Hà Nội ngày 1-5-2006

Tác giả

CHƯƠNG 1

NHỮNG KHÁI NIỆM CƠ BẢN VÀ PHƯƠNG PHÁP LUẬN CỦA KINH TẾ LƯỢNG

§1. KHÁI NIỆM CƠ BẢN CỦA KINH TẾ LƯỢNG

- Thuật ngữ tiếng Anh là: "Econometrics" có nghĩa là đo lường kinh tế. Thuật ngữ này do giáo sư người Na Uy A. K. Ragnar Frisch cùng J. Tinbergen sử dụng vào khoảng năm 1930.

- Phạm vi của kinh tế lượng được được hiểu thông qua một số định nghĩa sau:

+ Kinh tế lượng bao gồm việc áp dụng thống kê toán cho các số liệu kinh tế để củng cố về mặt thực nghiệm cho các mô hình.

+ Kinh tế lượng là một môn khoa học trong đó áp dụng các công cụ lý thuyết thống kê toán học và suy đoán thống kê để phân tích các vấn đề kinh tế.

- Kinh tế lượng là sự kết hợp các lý thuyết kinh tế, kinh tế toán, thống kê kinh tế, thống kê toán nhưng nó vẫn là một môn độc lập do có các lý do sau đây:

+ Các lý thuyết kinh tế thường nêu ra các giả thuyết hay các giả thiết. Phần lớn các giả thiết này nói về chất. Kinh tế lượng sẽ cho chúng ta ước lượng bằng số về chất này

+ Kinh tế toán thường trình bày lý thuyết kinh tế dưới dạng toán học (phương trình hoặc bất phương trình). Kinh tế lượng thường sử dụng phương trình toán học do các nhà kinh tế toán đề xuất và đặt các phương trình dưới dạng phù hợp để kiểm định bằng thực nghiệm.

+ Thống kê kinh tế chủ yếu liên quan đến việc thu thập, xử lý và trình bày các số liệu. Những số liệu này là số liệu thô đối với kinh tế lượng.

+ Các số liệu thu thập được đều là phi thực nghiệm, chứa sai số phép đo. Kinh tế lượng dùng các công cụ, phương pháp của thống kê toán để tìm ra bản chất của các số liệu thống kê.

§2. PHƯƠNG PHÁP LUẬN CỦA KINH TẾ LƯỢNG

Nêu ra các giả thuyết về các mối quan hệ giữa các biến kinh tế:

Bước 1: Thiết lập các mô hình toán học để mô tả mối quan hệ giữa các biến với nhau.

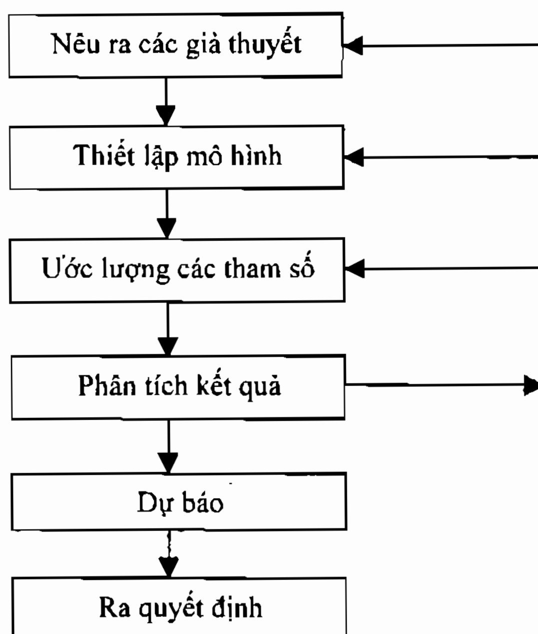
Bước 2: Ước lượng các tham số: nhằm nhận được số đo về mức ảnh hưởng của các biến với các số liệu hiện có. Các ước lượng này là kiểm định thực nghiệm cho lý thuyết kinh tế.

Bước 3: Phân tích kết quả: dựa trên lý thuyết kinh tế để phân tích, đánh giá kết quả nhận được, xem xét kết quả nhận được có phù hợp với lý thuyết kinh tế không. Sau đó kiểm định các giả thuyết thống kê về các ước lượng nhận được. Trong trường hợp không phù hợp với lý thuyết kinh tế cần đưa một mô hình đúng.

Bước 4: Dự báo: nếu mô hình phù hợp với lý thuyết kinh tế thì có thể sử dụng mô hình dự báo. Có hai dự báo là: dự báo giá trị trung bình hoặc dự báo giá trị cá biệt

Bước 5: Ra quyết định: sử dụng mô hình đề ra chính sách hợp lý

Nhìn vào sơ đồ dưới đây ta thấy rõ được phương pháp:



§3. MỘT SỐ KIẾN THỨC XÁC SUẤT THỐNG KÊ CẦN DÙNG

3.1. Đại lượng ngẫu nhiên và luật phân phối xác suất

3.1.1. Đại lượng ngẫu nhiên (hay biến ngẫu nhiên)

*) Các định nghĩa:

- Phép thử: mỗi sự thực hiện của tổ hợp các điều kiện nào đó gọi là một phép thử. Phép thử được gọi là ngẫu nhiên nếu không biết trước được kết cục của nó.

Ví dụ trong trường hợp gieo một con xúc sắc đồng chất trên một mặt phẳng. Phép thử này có 6 kết cục xảy ra: xuất hiện mặt 1 chấm, mặt 2 chấm, ..., mặt 6 chấm.

- Sự kiện: khái niệm thường gặp trong lý thuyết xác suất đó là sự kiện. Sự kiện được hiểu như là một sự việc, một hiện tượng nào đó của cuộc sống tự nhiên và xã hội.

- Sự kiện ngẫu nhiên: sự kiện có thể xảy ra hoặc không xảy ra được gọi là sự kiện ngẫu nhiên.

Ví dụ trong trường hợp gieo một con xúc sắc đồng chất trên một mặt phẳng. Phép thử này có 6 kết cục xảy ra: xuất hiện mặt 1 chấm, mặt 2 chấm, ..., mặt 6 chấm.

- Tần suất tương đối của sự kiện ngẫu nhiên: là tỷ số giữa số lần xuất hiện sự kiện m với số phép thử n tiến hành.

- Xác suất: trong nhiều hiện tượng hàng loạt khi thực hiện nhiều lần cùng một phép thử, ta thấy tần suất xuất hiện của sự kiện A nào đó không chênh nhiều với một con số đặc trưng cho khả năng xuất hiện A. Số đó được gọi là xác suất xuất hiện A và được ký hiệu là $P(A) = p$. Xác suất của sự kiện A là trị số ổn định của tần suất khi số phép thử tăng lên vô hạn.Vậy viết $P(A) = p$ thì p có nghĩa là xác suất xảy ra sự kiện A bằng p.

- Đại lượng ngẫu nhiên: kết quả của các phép thử được đặc trưng bởi các đại lượng biến thiên gọi là các đại lượng ngẫu nhiên.

*) Phân loại đại lượng ngẫu nhiên

- Đại lượng ngẫu nhiên rời rạc: đại lượng ngẫu nhiên được gọi là rời rạc nếu tập giá trị của nó là một tập hữu hạn hoặc vô hạn đếm được các phần tử.

- Đại lượng ngẫu nhiên liên tục: đại lượng ngẫu nhiên được gọi là liên tục

nếu tập giá trị của nó lấp kín một khoảng trên trục số.

3.1.2. Luật phân bố xác suất

- Định nghĩa luật phân bố xác suất: là một quy tắc nào đó cho phép tìm các xác suất của tất cả các giá trị có thể của đại lượng ngẫu nhiên.

- Đối với đại lượng ngẫu nhiên rời rạc, mỗi giá trị của nó được gắn với một xác suất đặc trưng cho khả năng đại lượng ngẫu nhiên nhận giá trị đó.

Ta có: $p_i = P(X = x_i)$, trong đó X là đại lượng ngẫu nhiên, x_i là giá trị của biến ngẫu nhiên X. Với loại đại lượng này luật phân bố xác suất cho dưới dạng bảng.

- Với bảng phân phối xác suất chưa biểu diễn được đại lượng ngẫu nhiên liên tục.Với loại đại lượng này dùng hàm phân phối xác suất.

*) Bảng phân phối xác suất

Bảng phân phối của đại lượng ngẫu nhiên X thể hiện trong bảng 1.

Bảng 1

$X = x$	x_1 x_2 ..... x_n ....
$p(x)$	p_1 p_2 ......p_n.........

trong đó x_1, x_2, ... x_n ... là các giá trị của X còn $p_n = p(x_n) = P(X = x_n)$ là các xác suất để đại lượng ngẫu nhiên X lần lượt nhận giá trị x_1, x_2, ... x_n ... $p_i \geq 0$; $\sum_{i=1}^{\infty} p_i = 1$.

*) Hàm phân phối xác suất

Bảng phân phối xác suất có một hạn chế cơ bản là chưa đủ tổng quát để đặc trưng cho một đại lượng ngẫu nhiên tuỳ ý, nhất là trong trường hợp đại lượng ngẫu nhiên liên tục. Vì vậy người ta đưa ra khái niệm hàm phân phối xác suất.

Hàm phân phối xác suất của đại lượng ngẫu nhiên X, ký hiệu là $F(x)$, được xác định như sau:

$$F(x) = P(X < x), x \in \boldsymbol{R}. \qquad (1.1)$$

Tính chất của hàm phân phối xác suất:

(i) $0 \le F(x) \le 1$

(ii) $F(x)$ là một hàm đồng biến

(iii) $P(\alpha \le X < \beta) = F(\beta) - F(\alpha)$

(iv) $F(+\infty) = 1; F(-\infty) = 0$

*) Hàm mật độ xác suất

Hàm phân phối xác suất còn hạn chế là không cho biết rõ phân phối xác suất ở lân cận một điểm nào đó trên trục số. Vì vậy đối với các đại lượng ngẫu nhiên liên tục có $F(x)$ khả vi, người ta đưa ra khái niệm hàm mật độ xác suất.

Hàm mật độ xác suất của đại lượng ngẫu nhiên X, ký hiệu là $f(x)$, có hàm phân phối xác suất $F(x)$ khả vi (trừ ở một số hữu hạn điểm gián đoạn bị chặn), được xác định bằng:

$$f(x) = F'(x) \quad (1.2)$$

Hay $$F(x) = \int_{-\infty}^{x} f(t)dt \quad (1.3)$$

Theo tính chất của hàm phân phối xác suất và theo công thức (1.1) thì:

$$P(\alpha \le X < \beta) = \int_{\alpha}^{\beta} f(x)dx \quad (1.4)$$

Tính chất của hàm mật độ xác suất:

(i) $f(x) \ge 0 \forall x$

(ii) $\int_{-\infty}^{+\infty} f(x)dx = 1$

3.1.3. Các thông số đặc trưng của đại lượng ngẫu nhiên

*) Kỳ vọng

Kỳ vọng của đại lượng ngẫu nhiên X, ký hiệu là EX được định nghĩa như sau:

- Nếu X là đại lượng ngẫu nhiên rời rạc có hàm xác suất:

$$p(x_i) = p_i,\ i = 1, 2 \ldots \text{ thì } EX = \sum_{\forall i} x_i p_i \quad (1.5)$$

- Nếu X là đại lượng ngẫu nhiên liên tục có hàm mật độ xác suất

$$f(x),\ x \in \mathbf{R} \text{ thì } EX = \int_{-\infty}^{+\infty} x f(x)dx \quad (1.6)$$

Tính chất của kỳ vọng

(i) $Ec = c$ trong đó c là hằng số;

(ii) $E(cX) = cEX$;

(iii) $E(X+Y) = EX+EY$;

(iv) Nếu X, Y là độc lập thì $E(XY) = EX.EY$

(v) Nếu $Y = \varphi(X)$, thì tuỳ vào X là liên tục hay rời rạc ta có:

$EY = \sum_i \varphi(x_i)p_i$ *hoặc* $EY = \int_{-\infty}^{+\infty} \varphi(x)f(x)dx$ *với p(x), f(x) là hàm xác suất và hàm mật độ xác suất.*

*) Phương sai

Phương sai của đại lượng ngẫu nhiên *X*, ký hiệu là *DX*, được định nghĩa như sau:

$$DX = E[(X - EX)^2] \tag{1.7}$$

- Đại lượng ngẫu nhiên *X* rời rạc thì

$$DX = \sum_{\forall i}(x_i - EX)^2 p_i = \sum_{\forall i} x_i^2 p_i - \left(\sum_{\forall i} x_i p_i\right)^2 \tag{1.8}$$

- Đại lượng ngẫu nhiên liên tục

$$DX = \int_{-\infty}^{+\infty}(x_i - EX)^2 f(x)dx = \int_{-\infty}^{+\infty} x^2 f(x)dx - \left(\int_{-\infty}^{+\infty} xf(x)dx\right)^2 \tag{1.9}$$

Độ lệch chuẩn $\sigma(X) = \sqrt{DX}$

Tính chất của phương sai:

(i) $Dc = 0$

(ii) $D(cX) = c^2 DX$

(iii) Nếu X,Y là độc lập thì $D(X + Y) = DX + DY$

hay $\sigma(X+Y) = \sqrt{\sigma^2(X) + \sigma^2(Y)}$

*) Phân phối chuẩn

Đại lượng ngẫu nhiên *X* được gọi là tuân theo luật phân phối chuẩn, ký hiệu là $X \sim N(a,\sigma^2)$, nếu hàm mật độ của nó có dạng

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}} ; x \in \mathbf{R} \quad (1.10)$$

trong đó a, σ^2 lần lượt là kỳ vọng và phương sai.

*) Tương quan giữa hai đại lượng ngẫu nhiên

Tương quan giữa hai đại lượng ngẫu nhiên X và Y, ký hiệu là $\text{cov}(X,Y)$ được tính như sau:

$$\text{cov}(X,Y) = E[(X - EX)(Y - EY)] \quad (1.11)$$

Nếu $\text{cov}(X,Y) = 0$ thì hai đại lượng X và Y không tương quan với nhau.

3.2. Mẫu ngẫu nhiên và các đặc trưng mẫu

3.2.1. Mẫu ngẫu nhiên từ một tập nền

Ta gọi mẫu ngẫu nhiên kích thước n từ tập nền có đại lượng ngẫu nhiên gốc X là một tập các biến $X_1, X_2, ..., X_n$ thoả mãn điều kiện:

+) Độc lập thống kê,

+) Có cùng phân phối với X

3.2.2. Các đặc trưng mẫu

- Trung bình mẫu:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad (1.12)$$

- Phương sai mẫu (có chệch):

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{X})^2 \quad (1.13)$$

- Phương sai mẫu hiệu chỉnh (không chệch):

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{X})^2 \quad (1.14)$$

- Độ lệch chuẩn của mẫu :

$$S = \sqrt{S^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{X})} \quad (1.15)$$

- Mômen bậc k của mẫu:

$$M_k = \frac{1}{n}\sum_{i=1}^{n} x_i^k \quad (1.16)$$

- Mômen trung tâm bậc k của mẫu:

$$S_k = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{X})^k \qquad (1.17)$$

Định lý 1:

Nếu đại lượng ngẫu nhiên gốc X tuân theo luật phân phối chuẩn $X \sim N(a,\sigma^2)$ thì $\overline{X}$ và S^2 độc lập với nhau và khi đó:

a) $\overline{X}$ hội tụ theo xác suất tới a ký hiệu $\overline{X}\xrightarrow{xs}a$, nghĩa là:

$$\lim p\left(\left|\overline{X} - a\right| < \varepsilon\right) = 1$$

b) $S^2 \xrightarrow{xs} \sigma^2 \qquad s^2 \xrightarrow{xs} \sigma^2$

c) $E\overline{X} = a, \qquad E(S^2) = \frac{n-1}{n}\sigma^2, \qquad E(s^2) = \sigma^2$

d) $M_k \xrightarrow{xs} E(X^k), \qquad S_k \xrightarrow{xs} M(X-a)^k$

3.3. Thống kê

3.3.1. Định nghĩa

- Một hàm $Y = g(X_1, X_2, ..., X_n)$ phụ thuộc vào tập giá trị của mẫu ngẫu nhiên được gọi là một thống kê.

- Do X_i nhận các giá trị tương ứng x_i nên hàm $g(x_1, x_2, ..., x_n)$ cũng được gọi là thống kê.

3.3.2. Các thí dụ về thống kê

*) $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i = g(x_1, x_2, ..., x_n)$

*) $S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{X})^2 = g(x_1, x_2, ..., x_n)$

*) $t = \frac{\overline{X} - a}{s}\sqrt{n} = g(x_1, x_2, ..., x_n)$

*) $\chi^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{X})^2}{\sigma^2} = g(x_1, x_2, ..., x_n)$

*) *Với mẫu 1:* $x_1, x_2, \ldots, x_m$

với mẫu 2: $y_1, y_2, \ldots, y_n$

thì $F = \dfrac{s_1^2}{s_2^2}$ *với* $s_1^2 \geq s_2^2$ *là một thống kê.*

3.3.3. Các định lý cần dùng

Định lý 2:

Nếu $X \sim N(a, \sigma^2)$ *thì các thống kê*

$$\chi^2 = \frac{nS^2}{\sigma^2} \qquad \chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

tuân theo phân phối "khi bình phương" với số bậc tự do là $m = n - 1$.

Định lý 3:

$X \sim N(a, \sigma^2)$ *thì thống kê*

$$t = \frac{\overline{X} - a}{\sigma}\sqrt{n} \qquad (a = const)$$

tuân theo phân phối Student với số bậc tự do là $m = n - 1$.

Định lý 4:

Giả sử có hai mẫu $x_1, x_2, \ldots, x_m$ *và* $y_1, y_2, \ldots, y_n$ *lấy ra từ hai tập nền có các đại lượng ngẫu nhiên tương ứng là* X, Y.

Nếu $X \sim N(a, \sigma^2), Y \sim N(a, \sigma^2)$ *thì thống kê* $F = \dfrac{s_1^2}{s_2^2}$ *với* $s_1^2 \geq s_2^2$ *tuân theo luật phân bố Fisher với bậc tự do của tử là* $m_1 = m-1$; $m_2 = n-1$

3.4. Ước lượng thống kê các tham số

3.4.1. Ước lượng điểm

*) Ước lượng tham số

Giả sử ta cần nghiên cứu một tập gốc có đại lượng ngẫu nhiên gốc X và có hàm phân phối xác suất đã biết nhưng chưa biết tham số θ nào đó, ta phải xác định giá trị của θ dựa trên các thông tin thu được từ một mẫu quan sát $x_1, x_2, \ldots, x_n$ của X. Giá trị tìm được trong quá trình ấy, ký hiệu là $\hat{\theta}$ sẽ được gọi là ước lượng của θ. Do $\hat{\theta}$ khi tính toán được sẽ là một trị số nên nó được gọi là ước lượng điểm.

*) Các tính chất của ước lượng điểm

a. Ước lượng không chệch: thống kê $\hat{\theta}$ được gọi là ước lượng không chệch của θ nếu $E\hat{\theta} = \theta$ .Ta thấy rằng $E(\hat{\theta} - \theta) = 0$, tức là trung bình độ lệch của ước lượng so với giá trị thật bằng 0. Nếu độ lệch có trung bình khác 0, ta có ước lượng chệch..Ta thấy rằng:

- Trung bình mẫu là ước lượng không chệch của kỳ vọng

- Phương sai mẫu hiệu chỉnh là ước lượng không chệch của phương sai.

- Phương sai tính theo công thức $\frac{1}{n-1}\sum_{i=1}^{n}(x_i - a)^2$ là ước lượng không chệch của phương sai, trong khi đó S^2 là ước lượng chệch.

b. Ước lượng vững: thống kê $\hat{\theta}$ được gọi là ước lượng vững của θ nếu $\hat{\theta}(x_1, x_2, ..., x_n) \xrightarrow[n\to\infty]{XS} \theta$. Ta thấy rằng nếu $\hat{\theta}$ là ước lượng tiệm cận không chệch của θ, tức là $\lim E\hat{\theta} = \theta; \lim D\hat{\theta} = 0$ khi $n \to \infty$ thì $\hat{\theta}$ là ước lượng vững. Rõ ràng $\overline{X}$, $S^2(s^2)$ là ước lượng vững của EX và DX.

c. Ước lượng hiệu quả: thống kê $\hat{\theta}$ là ước lượng hiệu quả của θ nếu nó là ước lượng không chệch có phương sai bé nhất.

Người ta chứng minh được rằng nếu $\hat{\theta}$ là ước lượng hiệu quả của θ thì phương sai của nó là:

$$\frac{1}{nE\left[\frac{\partial \ln f(x,\theta)}{\partial \theta}\right]^2} \quad (1.18)$$

3.4.2. Ước lượng khoảng

Giả sử tìm được khoảng $[g_{min}, g_{max}]$ chứa θ, mà $P(g_{min} < \theta < g_{max}) = 1 - \alpha$.

Khi đó $[g_{min}, g_{max}]$ là ước lượng khoảng của θ, $1 - \alpha$ là xác suất để $\theta \in [g_{min}, g_{max}]$ gọi là độ tin cậy.

Xét khoảng tin cậy cho kỳ vọng $a = EX$.

Ta chọn ước lượng của a là $\overline{x}$, của σ^2 là s^2.

Theo định lý 3 thì biến ngẫu nhiên $t = \frac{\overline{x} - a}{s}\sqrt{n}$ tuân theo luật phân phối Student với bậc tự do là $m = n - 1$. Vì vậy khi đã cho trước $0 < \alpha < 1$, tra bảng

Student ta tìm được số $t_{\alpha,m}$ sao cho:

$$P(|t| > t_{\alpha,m}) = \alpha \qquad (1.19)$$

Từ (1.19) ta có $P\left(\left|\frac{\overline{x}-a}{s}\sqrt{n}\right| > t_{\alpha,m}\right) = 1-\alpha$. Giải bất phương trình sau theo a ta được:

$$\overline{x} - t_{\alpha,m}\frac{s}{\sqrt{n}} < a < \overline{x} + t_{\alpha,m}\frac{s}{\sqrt{n}}$$

Do vậy: $P\left(\overline{x} - t_{\alpha,m}\frac{s}{\sqrt{n}} < a < \overline{x} + t_{\alpha,m}\frac{s}{\sqrt{n}}\right) = 1-\alpha$

⇒ Khoảng tin cậy của $a = EX$ với độ tin cậy $1 - \alpha$ là:

$$\left(\overline{x} - t_{\alpha,m}\frac{s}{\sqrt{n}}; \overline{x} + t_{\alpha,m}\frac{s}{\sqrt{n}}\right)$$

3.5. Kiểm định giả thuyết thống kê

3.5.1. Khái niệm về giả thuyết, đối thuyết

- Trong thống kê chúng ta xuất phát từ một mẫu $x_1, x_2, ..., x_n$ chọn từ một tập nền có hàm phân phối $F(x, \theta)$, nhưng chưa biết tham số θ. Nếu tham số θ chưa biết và giả thuyết θ bằng giá trị cụ thể θ_0 được đưa ra thì ta nói rằng có một giả thuyết đơn. Giả thuyết được đưa ra được gọi là giả thuyết gốc, ký hiệu là H_0. Các giả thuyết khác với giả thuyết gốc là đối thuyết ký hiệu là H_1

- Ví dụ: khi nghiên cứu thu nhập của cư dân một thành phố nào đó, ta có thể đưa ra nhiều giả thuyết khác nhau:

+) Thu nhập của cư dân tuân theo luật phân phối chuẩn (H_0), hoặc không tuân theo luật đó (H_1).

+) Thu nhập trung bình là 50 triệu đồng (H_0) với nhiều dạng đối thuyết khác nhau: thí dụ $\neq 50$, > 50...

3.5.2. Các bước kiểm định giả thuyết

Với bài toán kiểm định giả thuyết, ta có cặp giả thuyết H_0 và đối thuyết H_1. Cần có quyết định bác bỏ hay công nhận giả thuyết H_0 một cách hợp lý dựa trên những thông tin có trong mẫu $x_1, x_2, ..., x_n$ lấy từ tập gốc có hàm phân phối $F(x)$ và biến gốc.

Bước 1: Chọn một thống kê có liên quan đến H_0: $g(x_1, x_2, ..., x_n)$

Bước 2: Dựa vào các định lý thống kê ở trên để biết thống kê đó tuân theo luật phân phối nào.

Bước 3: Chọn mức ý nghĩa α. Tra bảng tương ứng ta sẽ được g_α.

Bước 4: Dựa vào mẫu tính $\hat{g}$

- Nếu $\hat{g} > g_\alpha \Rightarrow$ bác bỏ H_0.

- Nếu $\hat{g} \leq g_\alpha \Rightarrow$ công nhận H_0.

CHƯƠNG 2

MÔ HÌNH HỒI QUY HAI BIẾN, ƯỚC LƯỢNG VÀ KIỂM ĐỊNH GIẢ THIẾT

§1. PHÂN TÍCH HỒI QUY

- Phân tích hồi quy là một phương pháp phân tích thống kê nghiên cứu mối liên hệ phụ thuộc của một biến phụ thuộc (hay là biến được giải thích) với một hay nhiều biến độc lập (hay biến giải thích). Nó cũng được vận dụng để đánh giá hiệu quả tác động của biến độc lập với biến phụ thuộc.

- Xét các ví dụ :

+ Khi nghiên cứu nhu cầu về một loại hàng hoá thì thấy rằng nhu cầu về loại hàng hoá đó phụ thuộc vào giá của hàng hoá đó, phụ thuộc vào thu nhập của người tiêu dùng và giá của các loại hàng hoá đang cạnh tranh với loại hàng hoá này.

Ở đây nhu cầu của loại hàng hoá đó là biến phụ thuộc, còn giá bản thân của hàng hóa đó, thu nhập của người tiêu dùng, giá của các loại hàng hoá cạnh tranh với hàng hoá đó chính là các biến độc lập.

+ Giá nhà ở hiện hành phụ thuộc vào các yếu tố như: diện tích sử dụng, vị trí, chất lượng xây dựng của ngôi nhà đó và còn phụ thuộc vào nhiều yếu tố nữa như giá của ngôi nhà đó những năm trước... Ở đây giá nhà ở hiện hành là biến phụ thuộc, còn các yếu tố: vị trí, chất lượng, giá cả những năm trước của ngôi nhà... đó là những biến phụ thuộc.

- Các ký hiệu: Y - biến phụ thuộc hay biến được giải thích;

X_i - biến độc lập hay biến giải thích.

- Phân tích hồi quy sẽ giải quyết các vấn đề sau:

+ Ước lượng giá trị trung bình của biến phụ thuộc với giá trị của các biến độc lập.

+ Kiểm định giả thiết về bản chất của sự phụ thuộc.

+ Dự đoán giá trị trung bình của biến phụ thuộc khi biết giá trị của biến

độc lập.

+ Tổng hợp các vấn đề trên lại.

1.1. Quan hệ thống kê và quan hệ hàm số

Trong phân tích hồi quy, chúng ta phải phân biệt hai quan hệ là quan hệ thống kê và quan hệ hàm số để tiện khi sử dụng.

1.1.1. Quan hệ thống kê

- Ta thấy rằng biến được giải thích phụ thuộc vào các biến giải thích. Các biến giải thích, giá trị của chúng đã biết, còn biến được giải thích là đại lượng ngẫu nhiên có phân bố xác suất, ta phải tính toán và dự đoán chúng.

- Biến phụ thuộc là đại lượng ngẫu nhiên và có rất nhiều nhân tố tác động đến nó mà trong mô hình ta không thể đề cập ra được, ứng với mỗi giá trị khác nhau của biến độc lập ta lại có một biến phụ thuộc khác.

- Quan hệ thống kê biểu thị khái quát trung bình và chỉ biểu hiện thông qua nhiều quan sát.

- Năng suất một loại lúa trên một hecta (biến phụ thuộc) phụ thuộc vào nhiệt độ, lượng mưa, độ chiếu sáng, lượng phân bón... (là các biến độc lập). Năng suất là một biến phụ thuộc, là đại lượng ngẫu nhiên không thể dự báo một cách chính xác do có sai số trong các phép đo, và còn rất nhiều yếu tố ảnh hưởng đến nữa mà ta không thể tách bỏ được. Quan hệ giữa năng suất lúa vào các đại lượng khác là quan hệ thống kê.

1.1.2. Quan hệ hàm số

- Trong quan hệ hàm số, các biến không phải là ngẫu nhiên, ứng với mỗi giá trị của biến độc lập thì có một biến phụ thuộc. Phân tích hồi quy thường không để ý tới các quan hệ hàm số.

- Định luật Ohm quy định sự phụ thuộc hàm số giữa hiệu điện thế với điện trở và cường độ dòng điện của nó, không phụ thuộc vào độ dài và tiết diện của dây dẫn.

1.2. Hàm hồi quy và quan hệ nhân quả

Phân tích hồi quy nghiên cứu quan hệ giữa một biến phụ thuộc với nhiều biến độc lập khác. Điều này không đòi hỏi giữa biến độc lập và biến phụ thuộc

có mối quan hệ nhân quả. Nếu có sự tồn tại quan hệ nhân quả này thì nó phải được xác lập dựa trên các lý thuyết kinh tế khác.

Chẳng hạn ta có dự đoán năng suất của một loại lúa trên một hecta dựa vào lượng mưa, nhưng ngược lại thì không đúng vì không thể dự đoán lượng mưa dựa vào năng suất, do năng suất còn phụ thuộc vào nhiều yếu tố khác.

1.3. Hồi quy và tương quan

- Hồi quy và tương quan khác nhau về mục đích và kỹ thuật.

- Phân tích hồi quy: ước lượng hoặc dự báo một biến trên cơ sở giá trị đã cho của các biến khác, biến phụ thuộc là đại lượng ngẫu nhiên. Các biến giải thích giá trị của chúng đã biết.

- Phân tích tương quan đo mức độ kết hợp tuyến tính giữa hai biến, không có sự phân biệt giữa các biến, chúng có tính chất đối xứng.

1.4. Nguồn số liệu cho phân tích hồi quy

Việc thành công hay thất bại của bất kỳ một sự phân tích kinh tế nào đều phụ thuộc vào việc sử dụng số liệu cũng như phương pháp xử lý số liệu đó.

Vậy nên việc tìm hiểu về nguồn gốc, bản chất cũng như ưu, nhược điểm của số liệu cũng khá quan trọng.

1.4.1. Các loại số liệu

Có 3 loại số liệu.

a. Số liệu theo thời gian:

- Số liệu theo thời gian là số liệu thu thập được trong một thời gian cố định chẳng hạn như số liệu về GNP, số lượng người thất nghiệp, lượng cung về tiền... Có số liệu thu thập theo ngày, tháng, năm...

- Các số liệu này có thể đo bằng con số như giá cả, thu nhập. Nhưng cũng có số liệu không đo được bằng con số mà chúng thường là chất lượng như: đã kết hôn hay chưa, có việc làm hay thất nghiệp.

b. Số liệu chéo

- Số liệu chéo là số liệu về một hay nhiều biến được thu thập tại một thời điểm ở nhiều địa phương, đơn vị khác nhau. Chẳng hạn như số liệu điều tra về dân số của Việt Nam vào 01//01/2006...

c. Số liệu hỗn hợp theo thời gian và không gian: các số liệu về giá vàng ở Hà Nội, Thành phố Hồ Chí Minh, Hải Phòng...

1.4.2. Nguồn gốc các số liệu

- Các số liệu có thể do các cơ quan nhà nước, các tổ chức quốc tế, các công ty tư nhân hay các cá nhân thu thập. Chúng có thể là số liệu thực nghiệm hay không thực nghiệm.

- Các số liệu thực nghiệm thường được thu thập trong khoa học tự nhiên, muốn thu thập các số liệu ảnh hưởng tới đối tượng nghiên cứu phải giữ nguyên các yếu tố khác. Ngược lại trong khoa học xã hội các số liệu không phải do thực nghiệm mà có.

1.4.3. Nguyên nhân gây nên nhược điểm của các số liệu

- Hầu hết các số liệu trong khoa học xã hội đều không phải là số liệu thực nghiệm. Do vậy có sai số quan sát hoặc bỏ sót quan sát, thậm chí cả hai.

- Các số liệu thu thập bằng thực nghiệm cũng có sai số phép đo.

- Trong các cuộc điều tra không biết rõ câu hỏi.

- Khó khăn trong việc so sánh giữa các kết quả điều tra do các mẫu điều tra có kích thước khác nhau...

§2. HỒI QUY HAI BIẾN

Trường hợp đơn giản nhất của phân tích hồi quy, người ta nghiên cứu mối liên hệ giữa hai chỉ tiêu , mà một chỉ tiêu được xem như yếu tố độc lập (trị số của nó ký hiệu là X) còn chỉ tiêu kia là biến phụ thuộc (trị số của nó ký hiệu là Y).

Một trong các bài toán đầu tiên của phân tích hồi quy là xác định dạng hàm số này, tức là tìm phương trình hồi quy đảm bảo các mối liên hệ được nghiên cứu.

2.1. Phương trình hồi quy

Phương trình hồi quy là bộ phận cấu thành quan trọng nhất của các mô hình hồi quy. Việc lựa chọn và tính toán đúng đắn phương trình này là một bước quan trọng nhất trong việc lập mô hình hồi quy.

Giả sử X là biến độc lập dùng để xác định biến phụ thuộc Y, $Y = f(X)$.

Ta xét đơn giản ở đây biến Y phụ thuộc tuyến tính vào $X \Rightarrow$ đường hồi quy là đường thẳng. Giả sử để xác định phương trình của đường thẳng hồi quy ta thực hiện n phép thử độc lập và thu được các cặp giá trị số tương ứng của các đinh X và Y là: (x_1,y_1), (x_2,y_2), ..., (x_n,y_n).

Do đường hồi quy là đường thẳng nên có dạng:

$$Y = aX + b \tag{2.1}$$

Có nhiều cách để xác định đường hồi quy. Tức là tìm các hệ số a,b. Ở đây ta dùng phương pháp BPCT để xác định đường hồi quy.

2.2. Phương pháp BPCT

Xét hàm tuyến tính một biến (2.1), với mỗi phép thử (x_i,y_i) $i = 1...n$, thay x_i vào phương trình (2.1).

Ta có:

$$Y_i = ax_i + b_i = 1...n$$

Lập các hiệu $(Y_i - y_i)$.

Theo phương pháp bình phương cực tiểu ta sẽ xác định các hệ số a, b sao cho tổng:

$$S = \sum_{i=1}^{n} (Y_i - y_i)^2 = \sum_{i=1}^{n} (ax_i + b - y_i)^2$$

đạt giá trị nhỏ nhất. Muốn vậy thì:

$$\begin{cases} \dfrac{\partial F}{\partial a} = 2\sum\limits_{i=1}^{n}(ax_i + b - y_i)x_i = 0 \\ \dfrac{\partial F}{\partial b} = 2\sum\limits_{i=1}^{n}(ax_i + b - y_i) = 0 \end{cases} \tag{2.2}$$

hay:

$$\begin{cases} a\sum\limits_{i=1}^{n} x_i^2 + b\sum\limits_{i=1}^{n} x_i = \sum\limits_{i=1}^{n} x_i y_i \\ a\sum\limits_{i=1}^{n} x_i + bn = \sum\limits_{i=1}^{n} y_i \end{cases} \tag{2.3}$$

Giải hệ phương trình (2.2) bằng phương pháp Crame ta có:

$$\left.\begin{aligned}\hat{a} &= \frac{n\sum_{i=1}^{n}x_iy_i - \sum_{i=1}^{n}x_i\sum_{i=1}^{n}y_i}{n\sum_{i=1}^{n}x_i^2 - (\sum_{i=1}^{n}x_i)^2} \\ \hat{b} &= \frac{\sum_{i=1}^{n}x_i^2\sum_{i=1}^{n}y_i - \sum_{i=1}^{n}x_i\sum_{i=1}^{n}x_iy_i}{n\sum_{i=1}^{n}x_i^2 - (\sum_{i=1}^{n}x_i)^2}\end{aligned}\right\} \quad (2.4)$$

Với các giá trị xác định $\hat{a},\hat{b}$ ta có phương trình hồi quy thực nghiệm sau:

$$\hat{y} = \hat{a}x + \hat{b} \quad (2.5)$$

§3. HỆ SỐ TƯƠNG QUAN - ĐỘ CHÍNH XÁC CỦA CÁC ƯỚC LƯỢNG BPCT

3.1. Hệ số tương quan

Giả sử có n cặp số liệu quan sát $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$

Và $$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad \bar{y} = \sum_{i=1}^{n} y_i \quad (2.6)$$

là các trung bình mẫu tương ứng của các biến X và Y.

Hệ số tương quan của mẫu điều tra được xác định như sau:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\left[\sum_{i=1}^{n}(x_i - \bar{x})^2\right]\left[\sum_{i=1}^{n}(y_i - \bar{y})^2\right]}} \quad (2.7)$$

hay:

$$r = \frac{n\sum_{i=1}^{n}x_iy_i - (\sum_{i=1}^{n}x_i)(\sum_{i=1}^{m}y_i)}{\sqrt{\left[n\sum_{i=1}^{n}x_i^2 - (\sum_{i=1}^{n}x_i)^2\right]\left[n\sum_{i=1}^{n}y_i^2 - (\sum_{i=1}^{n}y_i)^2\right]}} \quad (2.8)$$

Đặt: $$\left.\begin{aligned}S_X &= \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2} \\ S_Y &= \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2}\end{aligned}\right\} \quad (2.9)$$

sau đó thay vào (2.4) và (2.8) ta có:

$$r = \hat{a}\frac{S_X}{S_Y} \tag{2.10}$$

Có ba tính chất:

a) a − 1 ≤ r ≤ 1

b) Nếu $r = \pm 1$ thì các điểm (x_i, y_i) sẽ nằm trên một đường thẳng và ngược lại. Tức là khi các điểm (x_i, y_i) nằm trên một đường thẳng thì $r = \pm 1$

c) Vì $r = \hat{a}\frac{S_X}{S_Y}$ nên phương trình đường hồi quy thực nghiệm có dạng:

$$\hat{Y} = (r\frac{S_Y}{S_X})X + \hat{b} \tag{2.11}$$

Các số $S_Y \geq 0$, $S_X \geq 0$ nên nếu $r < 0$ thì ta có tương quan nghịch giữa Y và X, còn nếu $r > 0$, ta có tương quan thuận.

Chứng minh

a) Theo bất đẳng thức Bunhiacopski ta có:

$$\left(\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})\right)^2 \leq \sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2$$

nên

$$r^2 = \frac{\left(\sum\limits_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})\right)^2}{\left[\sum\limits_{i=1}^{n}(x_i - \bar{x})^2\right]\left[\sum\limits_{i=1}^{n}(y_i - \bar{y})^2\right]} \leq 1 \tag{2.12}$$

$$\Rightarrow -1 \geq r \leq 1$$

b) Dấu bằng xảy ra ở (2.12) khi $\frac{x_i - \bar{x}}{y_i - \bar{y}} = \frac{x_j - \bar{x}}{y_j - \bar{y}} \qquad i, j = 1,2,...,n$

ta thấy rằng đấy chính là phương trình đường thẳng. Điều ngược lại cũng rõ ràng đúng.

c) Nếu $r < 0$ thì hàm số ở vế trái của phương trình (2.11) là nghịch biến và $r > 0$ thì hàm nghịch biến do chúng là hàm tuyến tính bậc nhất .

3.2. Độ chính xác của các ước lượng BPCT

Ký hiệu:

$$X_i = x_i - \bar{x}, \quad Y_i = y_i - \bar{y}$$

σ^2 là phương sai lý thuyết chưa biết. Ta có:

$$\hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

- Phương sai của và độ lệch chuẩn của $\hat{a}$ là:

$$D\hat{a} = \frac{\hat{\sigma}^2}{\sum_{i=1}^{n} X_i^2} \quad (2.13) \quad \text{và} \quad Se(\hat{a}) = \sqrt{D\hat{a}} = \frac{\hat{\sigma}}{\sqrt{\sum_{i=1}^{n} X_i^2}} \quad (2.14)$$

- Phương sai và độ lệch chuẩn của $\hat{b}$ là:

$$D\hat{b} = \frac{\sum_{i=1}^{n} x_i^2}{\sum_{i=1}^{n} X_i^2}\hat{\sigma}^2 \quad (2.15) \quad \text{và} \quad Se(\hat{b}) = D\hat{b} = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{\sum_{i=1}^{n} X_i^2}}\hat{\sigma} \quad (2.16)$$

- Phương sai và độ lệch chuẩn của hàm số :

$$D\hat{Y} = \frac{1}{n-2}\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\bar{y}} \quad (2.17)$$

và

$$Se(\hat{Y}) = \sqrt{D\hat{Y}} = \sqrt{\frac{1}{n-2}\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\bar{y}}} \quad (2.18)$$

§4. PHƯƠNG TRÌNH ƯỚC LƯỢNG VÀ DỰ ĐOÁN

4.1. Phương trình ước lượng

Bằng phương pháp binh phương cực tiểu, ta tìm được đường hồi quy thực nghiệm có dạng :

$$\hat{Y} = \hat{a}X + \hat{b} \quad (2.19)$$

Đây thực chất là một ước lượng quan hệ phụ thuộc giữa X và Y. Do vậy có thể xem đó là một phương trình ước lượng

4.2. Một số dự đoán

4.2.1. Trường hợp nội suy

Trong nhiều bài toán ta xem phương trình tìm thấy $y = f(x)$ là một biểu hiện mối quan hệ tiềm ẩn có trong hệ thống, giữa hai tham số X và Y. Như vậy khi $X = x_0$ thì $Y = f(x_0)$

Phương trình đường hồi quy tuyến tính là: $f(x_0) = ax_0 + b$.

4.2.2. Trường hợp dự báo

Một dạng dự báo được xét với $X = T$ = thời gian. Dãy số liệu là dãy thời gian $(y_1, t_1), (y_2, t_2), \ldots, (y_m, t_m)$

§5. VÍ DỤ

Trong một thí nghiệm để khảo sát sự phụ thuộc giữa Y vào X, ta được các số liệu như bảng 2.1.

Lập phương trình hồi quy thực nghiệm:

Bảng 2.1

x_i	0	1	2	3	4	5	6
y_i	4,25	3	3	1,75	1,5	0,5	0,25

Áp dụng các công thức (2.4) ta tính được các hệ số $\hat{a}$ và $\hat{b}$

$$\hat{a} = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2} = \frac{7\times 24,25 - 21\times 14,25}{7\times 91 - 21\times 21} = \frac{-129,5}{196} = -0,66$$

$$\hat{b} = \frac{\sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i y_i}{n\sum_{i=1}^{n} x_i^n - (\sum_{i=1}^{n} x_i)^2} = \frac{91\times 14,25 - 21\times 24,25}{196} = \frac{787,5}{196} = 4,02$$

=> Phương trình hồi quy thực nghiệm là: $\hat{y} = -0,66x + 4,02$

Để tiện tính toán, ta lập bảng như bảng 2.2.

Bảng 2.2

x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
0	4,25	−3	2,215	-6,645	9	4,906
1	3,00	−2	0,965	-1,930	4	0,931
2	3,00	−1	0,965	-0,965	1	0,931
3	1,75	0	−0,285	0,000	0	0,081
4	1,50	1	−0,535	-0,535	1	0,286
5	0,50	2	−1,535	-3,070	4	2,356
6	0,25	3	−1,785	-5,355	9	3,185
$\Sigma x_i = 21$ $\bar{x}=3$	$\Sigma y_i = 14{,}25$ $\bar{y} = 2{,}035$			$\Sigma(x_i - \bar{x})(y_i - \bar{y}) = -18{,}5$	$\sum(x_i - \bar{x})^2 = 28$	$\Sigma(y_i - \bar{y})^2 = 12{,}676$

Theo công thức (2.7) hệ số tương quan của mẫu là:

$$r = \frac{n\sum\limits_{i=1}^{n} x_i y_i - (\sum\limits_{i=1}^{n} x_i)(\sum\limits_{i=1}^{n} y_i)}{\sqrt{\left[n\sum\limits_{i=1}^{n} x_i^2 - (\sum\limits_{i=1}^{n} x_i)^2\right]\left[n\sum\limits_{i=1}^{n} y_i^n - (\sum\limits_{i=1}^{n} y_i)^2\right]}}$$

Áp dụng công thức 2.11; 2.12; 2.13; 2.14; 2.15; 2.16; 2.17 để tính các ước lượng BPCT.

$$\hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

$$= \frac{1}{7-2}[(4,25 - 4,02)^2 + (3 - 3,36)^2 + (3 - 2,7)^2 +$$

$$+ (1,75 - 2,04)^2 + (1,5 - 1,38)^2 + (0,5 - 0,72)^2 + (0,25 - 0,06)^2]$$

$$= 0{,}0911$$

$$X_i = x_i - \bar{x}$$

$$\sum_{i=1}^{h} X_i^2 = (-3)^2 + (-2)^2 + (-1)^2 + 0^2 + 1^2 + 2^2 + 3^2 + 1 = 28$$

$$D\hat{a} = \frac{\sigma^2}{\sum_{i=1}^{n} X_i^2} = \frac{0,0911}{28} = 0,000325$$

$$Se(\hat{a}) = \sqrt{D\hat{a}} = 0,057$$

$$D\hat{b} = \frac{\sum_{i=0}^{n} x_i^2}{n\sum_{i=1}^{n} X_i^2}\sigma^2 = \frac{91}{7 \times 28}0,0911 = 0,0423$$

$$Se(\hat{b}) = \sqrt{D\hat{b}} = 0,2057$$

$$D\hat{y} = \frac{1}{n-2}\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\overline{y}} = \frac{0,0911}{2,035} = 0,0448$$

$$Se(\hat{y}) = \sqrt{D\hat{y}} = 0,2116$$

CHƯƠNG 3
HỒI QUY BỘI

§1. ĐẶT BÀI TOÁN

Trong chương 2 ta đã xét vấn đề hồi quy khi biến Y phụ thuộc vào một biến X, có thể coi đó là hồi quy đơn.

Trong chương này chúng ta xét trường hợp biến Y phụ thuộc vào hai biến trở lên gọi là hồi quy bội.

Bài toán tổng quát đặt ra như sau:

Giả sử cần nghiên cứu một đại lượng Y gọi là biến ra trong một hệ thống nào đó. Trong hệ thống đó một mặt Y phụ thuộc vào:

- Các biến số độc lập $X_1, X_2, \ldots, X_k$ gọi là các biến vào hay các nhân tố có thể điều khiển được.

- Mặt khác Y còn bị ảnh hưởng của các tác động ngẫu nhiên gọi là biến nhiễu hay sai số không thể điều khiển được.

Bài toán đặt ra là tìm quan hệ giữa Y và $X_1, X_2, \ldots, X_k$. Mối quan hệ giữa Y và $X_1, X_2, \ldots, X_k$ có dạng :

$$Y = f(X_1, X_2, \ldots, X_k; \theta_1, \theta_2, \ldots, \theta_m) + \xi, \tag{3.1}$$

trong đó dạng của hàm f đã biết, còn các tham số $\theta_1, \theta_2, \ldots, \theta_m$ là chưa biết, còn với $E\xi = 0$ và $D\xi = \sigma^2$, tức là biến ξ tuân theo phân phối chuẩn $\xi \sim N(0, \sigma^2)$.

Sự phụ thuộc giữa biến Y theo các biến $X_1, X_2, \ldots, X_k$ là tương đối phức tạp. Tuỳ từng bài toán đưa ra mà ta tìm được mô hình tương ứng.

Có nhiều mô hình hồi quy nhưng ở đây ta xét một số mô hình.

§2. MÔ HÌNH HỒI QUY TUYẾN TÍNH CỔ ĐIỂN

2.1. Mô hình hóa dưới dạng ma trận

Mô hình hồi quy tuyến tính cổ điển khẳng định Y phụ thuộc tuyến tính vào X_j ($j = 1, 2, \ldots, k$) Y là biểu thức bậc nhất của $X_1, X_2, \ldots, X_k$ và sai số ngẫu nhiên

hay chính là biến nhiễu của hệ thống.

Mô hình:

$$Y = \theta_0 + \theta_1 X_1 + \theta_2 X_2 + \ldots + \theta_k X_k + \xi$$

$$\xi \sim N(0, \sigma^2) \tag{3.2}$$

Trong đó θ_j $(j = 0, 1, \ldots, k)$ là các hệ số chưa biết.

Bây giờ ta tiến hành n quan sát đồng thời về $k + 1$ biến $X_1, X_2, \ldots, X_k$, và Y. Giá trị của quan sát tại phép thử thứ i là (x_i, y_i) Giả sử số liệu quan sát tuân theo mô hình sau:

$$\begin{aligned}
y_1 &= \theta_0 + \theta_1 x_{11} + \theta_2 x_{12} + \ldots + \theta_k x_{1k} + \xi_1 \\
y_2 &= \theta_0 + \theta_1 x_{21} + \theta_2 x_{22} + \ldots + \theta_k x_{2k} + \xi_2 \\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
y_n &= \theta_0 + \theta_1 x_{n1} + \theta_2 x_{n2} + \ldots + \theta_k x_{nk} + \xi_n
\end{aligned} \tag{3.3}$$

Mô hình (3.3) có các giả thiết sau:

+ Các ξ_j có kỳ vọng bằng 0, tức là $E(\xi_j) = 0$;

+ Không có sự tương quan giữa các ξ_i, $cov(\xi_i, \xi_j) = 0$

+ $D\xi_i = \sigma^2$

+ $\xi \sim N(0, \sigma^2)$

+ Biến độc lập này không là tổ hợp tuyến tính của các biến khác.

Mô hình (3.3) có thể viết dưới dạng ma trận sau:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \ldots & x_{1k} \\ 1 & x_{21} & x_{22} & \ldots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \ldots & x_{nk} \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_k \end{bmatrix} + \begin{bmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_n \end{bmatrix} \tag{3.4}$$

hay $\quad Y = X\theta + \xi$

trong đó:

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}; \quad X = \begin{bmatrix} 1 & x_{11} & x_{12} & \ldots & x_{1k} \\ 1 & x_{21} & x_{22} & \ldots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \ldots & x_{nk} \end{bmatrix}; \quad \theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_k \end{bmatrix}; \quad \xi = \begin{bmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_n \end{bmatrix}$$

X là ma trận thiết kế.

2.2. Ước lượng BPCT

Ta giả sử α_i là giá trị của θ_i ở phép thử thứ i. Khi đó giữa giá trị có được ở quan sát thứ i và giá trị tính được khi $\alpha_i = \theta_i$ sẽ lệch đi $y_i - \hat{y}_i$, với $i = 1, 2, ..., n$.

$$y_i - \hat{y}_i = y_i - (\alpha_0 + \alpha_1 x_{i1} + ... + \alpha_k x_{ik}) \tag{3.5}$$

Ở đây ta sẽ ước lượng các tham số của phương trình (3.5) bằng phương pháp bình phương cực tiểu

$$S(\alpha_0, \alpha_1, ..., \alpha_k) = \sum_{i=1}^{n} (y_i - \alpha_0 - \alpha_1 x_{i1} - ... - \alpha_k x_{ik})^2 = \sum_{i=1}^{n} \xi_i^2 \to \min \tag{3.6}$$

Ta sẽ viết (3.6) dưới dạng ma trận $S = (Y - XA)^T (Y - XA)$ với A là vector các hệ số α_i.

Đại lượng $\hat{\theta}_i$ chính là ước lượng bình phương cực tiểu của θ_i và $\hat{\xi}_i = y_i - \hat{y}_i = y_i - (\hat{\theta}_0 + \hat{\theta}_1 x_{i1} + ... + \hat{\theta}_k x_{ik})$ $i = 1, 2, .., n$ là ước lượng của sai số ngẫu nhiên ξ_i.

Phương trình $\hat{Y} = \hat{\theta}_0 + \hat{\theta}_1 X_1 + ... + \hat{\theta}_k X_k$ (3.7)

là phương trình hồi quy thực nghiệm.

Mệnh đề 3.1

Nếu ma trận thiết kế X không ngẫu nhiên có hạng k + 1 ≤ n thì ước lượng bình phương cực tiểu có dạng:

$$\hat{\theta} = (X^T X)^{-1} X^T Y \tag{3.8}$$

Khi đó:

$$\hat{Y} = X\hat{\theta} = X(X^T X)^{-1} X^T Y = HY \tag{3.9}$$

trong đó:

$$H = X(X^T X)^{-1} X^T \qquad H \text{ cấp } n.n \tag{3.10}$$

$$\tilde{\xi} = Y - \hat{Y} = (I_n - H)Y \tag{3.11}$$

Thỏa mãn:

$$X^T \hat{\xi} = 0 \quad \text{và} \quad \hat{Y}\hat{\xi} = 0, \quad (\hat{\theta}^T X^T \hat{\xi} = 0)$$

Tổng các phần dư:

$$\sum_{i=1}^{n} \hat{\xi}_i^2 = \hat{\xi}^T \hat{\xi} = Y^T Y - Y^T X \hat{\theta} \tag{3.12}$$

Chứng minh

Vì $S(\alpha)$ là hàm bậc 2 theo α nên $\hat{\theta}$ là ước lượng của θ có thể tìm được từ hệ phương trình sau:

$$\frac{\partial S}{\partial \alpha_j} = 0, j = 0..k$$

Hoặc tương đương:

$$\frac{\partial S}{\partial \alpha_0} = -2\sum_{i=1}^{n}(y_i - \alpha_0 - \alpha_1 x_{i1} - ... - \alpha_k x_{ik}) = 0$$

$$\frac{\partial S}{\partial \alpha_1} = -2\sum_{i=1}^{n}(y_i - \alpha_0 - \alpha_1 x_{i1} - ... - \alpha_k x_{ik})x_{i1} = 0$$

....

$$\frac{\partial S}{\partial \alpha_k} = -2\sum_{i=1}^{n}(y_i - \alpha_0 - \alpha_1 x_{i1} - ... - \alpha_k x_{ik})x_{ik} = 0$$

$$\Leftrightarrow \begin{cases} \sum_{i=1}^{n}(\alpha_0 + \alpha_0 x_{i1} + ... + \alpha_k x_{ik}) = \sum_{i=1}^{n} y_i \\ \alpha_0 \sum_{i=1}^{n} x_{i1} + \alpha_1 \sum_{i=1}^{n} x_{i1}^2 + ... + \alpha_k \sum_{i=1}^{n} x_{i1}x_{ik} = \sum_{i=1}^{n} y_i x_{i1} \\ .... \\ \alpha_0 \sum_{i=1}^{n} x_{ik} + \alpha_1 \sum_{i=1}^{n} x_{ik}x_{i1} ... + \alpha_k \sum_{i=1}^{n} x_{ik}^2 = \sum_{i=1}^{n} y_i x_{ik} \end{cases} \quad (3.13)$$

Ta thấy rằng nếu đặt $x_{i0} = 1$ với $i = 1, 2, ..., n$ ta có phương trình viết dưới dạng ma trận sau:

$$\begin{bmatrix} \sum_{i=1}^{n} x_{i0}^2 & \sum_{i=1}^{n} x_{i0}x_{i1} & ... & \sum_{i=0}^{n} x_{i0}x_{ik} \\ \sum_{i=1}^{n} x_{i1}x_{i0} & \sum_{i=1}^{n} x_{i1}^2 & ... & \sum_{i=1}^{n} x_{i1}x_{ik} \\ ... & ... & ... & ... \\ \sum_{i=1}^{n} x_{ik}x_{i0} & \sum_{i=1}^{n} x_{ik}x_{i1} & ... & \sum_{i=1}^{n} x_{ik}^2 \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_k \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} y_i x_{i0} \\ \sum_{i=1}^{n} y_i x_{i1} \\ \vdots \\ \sum_{i=1}^{n} y_i x_{ik} \end{bmatrix}$$

Hoặc dưới dạng ma trận

$$X^T X A = X^T Y \quad (3.14)$$

Vì $rank(X) = k + 1 \le n$ nên $X^T X$ là ma trận vuông cấp $k + 1$ nên có ma trận nghịch đảo là $(X^T X)^{-1}$.

Từ phương trình (3.14) ta có nghiệm:

$A = \hat{\theta} = (X^T X)^{-1} X^T Y$ => biểu thức (3.8) được chứng minh.

Ước lượng cho sai số ξ là $\hat{\xi}$

$$\hat{\xi} = Y - \hat{Y} = Y - X\hat{\theta}$$
$$= Y - X(X^T X)^{-1} X^T Y$$
$$= Y - HY = (I_n - H)Y$$

Với ma trận H vuông cấp n được xác định như trong biểu thức (3.10).

Để chứng minh $\hat{\theta}$ cực tiểu hóa $S(\alpha)$ và thỏa mãn đẳng thức (3.12) với chú ý rằng ma trận H có tính chất sau:

1. $(I - H)$ là đối xứng : $(I - H)^T = (I - H)$ (3.13)

2. $(I - H)^2 = (I - H)$ tức là $(I - H)$ là ma trận lũy đẳng.

3. $X^T (I - H) = X^T (I - X(X^T X)^{-1} X^T) = X^T - X^T = 0$

Do vậy nên

$$S(A) = (Y - XA)^T (Y - XA) = (Y - X\hat{\theta} + X\hat{\theta} - XA)^T (Y - X\hat{\theta} + X\hat{\theta} - XA)$$
$$= (Y - X\hat{\theta})^T (Y - X\hat{\theta}) + (\hat{\theta} - A)^T X^T X(\hat{\theta} - A) +$$
$$(\hat{\theta} - A)^T X^T (I - H)Y + Y^T (I - H)^T X(\hat{\theta} - A)$$
$$= (Y - X\hat{\theta})^T (Y - X\hat{\theta}) + (\hat{\theta} - A)^T X^T X(\hat{\theta} - A)$$
$$\ge (Y - X\hat{\theta})^T (Y - X\hat{\theta}) = S(\theta)$$

Dấu "=" xảy ra khi $\hat{\theta} = A$. Mà:

$$\sum_{i=1}^{n} \hat{\xi}_i^2 = S(\hat{\theta}) = (Y - X\hat{\theta})^T (Y - X\hat{\theta})$$
$$= Y^T (I - H)(I - H)Y = Y^T (I - H)Y$$
$$= Y^T Y - Y^T HY = Y^T Y - (Y^T X)\hat{\theta}$$

Biểu thức (3.12) được chứng minh.

Ví dụ 3.1: Để nghiên cứu sự phụ thuộc giữa doanh thu (Y) và chi phí sản xuất (X_1), chi phí tiếp thị (X_2), người ta điều tra ngẫu nhiên doanh thu của 12 công ty trong 12 thời kỳ, kết quả ta có bảng 3.1.

Bảng 3.1

X_0	X_1	X_2	Y
1	18	10	127
1	25	11	149
1	19	6	106
1	24	16	163
1	15	7	102
1	26	17	180
1	25	14	161
1	16	12	128
1	17	12	139
1	23	12	144
1	22	14	159
1	15	15	138

Giả sử chi phí này tuân theo mô hình tuyến tính cổ điển, khi đó:

$$y_i = \theta_0 + \theta_1 x_{i1} + \theta_2 x_{i2} + \xi_i; \quad i = 1,2,...,12$$

Ta sẽ ước lượng các hệ số hồi quy bằng ước lượng BPCT. Ta có:

$$X^T X = \begin{bmatrix} n & n\overline{x}_1 & n\overline{x}_2 \\ 0 & n\overline{x}_1^2 & n\overline{x}_2 \\ 0 & 0 & n\overline{x}_2^2 \end{bmatrix} = \begin{bmatrix} 12 & 245 & 146 \\ 0 & 5195 & 3055 \\ 0 & 0 & 1900 \end{bmatrix};$$

$$X^T Y = \begin{bmatrix} n\overline{y} \\ n\overline{x_1 y} \\ n\overline{x_2 y} \end{bmatrix} = \begin{bmatrix} 1696 \\ 35463 \\ 21409 \end{bmatrix} \Rightarrow$$

$$(X^T X)^{-1} = \begin{bmatrix} 2,439963 & -0,0883875 & -0,045374 \\ 0 & 0,006757 & -0,004040 \\ 0 & 0 & 0,010509 \end{bmatrix}$$

Theo mệnh đề 3.1 thì $\hat{\theta}=\begin{bmatrix}\hat{\theta}_0\\ \hat{\theta}_1\\ \hat{\theta}_2\end{bmatrix}=(X^TX)^{-1}X^TY=\begin{bmatrix}32,2777\\ 2,5057\\ 4,7587\end{bmatrix}$

Vậy dạng thực nghiệm của phương trình hồi quy tuyến tính là:

$$\hat{Y}=32,2777+2,5057X_1+4,7587X_2$$

Cũng theo mệnh đề 3.1 thì tổng bình phương các phần dư sẽ là:

$$\sum_{i=1}^{n}\hat{\xi}_i^2=\hat{\xi}^T\hat{\xi}=Y^TY-Y^TX\hat{\theta}$$

$$=245626-1696\times 32,2777-2,5057\times 35463-4,7587\times 21409$$

$$=144,3734$$

Ta có sai số bình phương trung bình $\frac{1}{12}\sum_{i=1}^{n}\hat{\xi}_i^2=\frac{144,3734}{12}\approx 12$ khá nhỏ. Như vậy mô hình khá phù hợp.

Bảng 3.2. tính các giá trị để tìm sai số $\hat{\xi}_i$ trực tiếp :

Bảng 3.2

i	y_i	$\hat{y}_i$	$\hat{\xi}_i$
1	127	124,9666	2,033
2	149	127,2659	1,734
3	106	108,4382	–2,438
4	163	168,5537	–5,554
5	102	103,1741	–1,174
6	180	178,3238	1,676
7	161	161,5420	–0,542
8	128	129,4733	–1,473
9	139	131,979	7,021
10	144	147,0132	–3,013
11	159	154,0249	4,975
12	138	141,2437	–3,244

Nếu tính trực tiếp ta có: $\sum_{i=1}^{n}\hat{\xi}_i^2=144,2298$.

2.2.1. Tính chất ước lượng bằng phương pháp bình phương cực tiểu

a) Ước lượng $\hat{\theta}$ là một ước lượng không chệch với:

$$E\hat{\theta} = \theta, \operatorname{cov}(\hat{\theta}) = \sigma^2 (X^T X)^{-1} \quad (3.18)$$

b) Sai số $\hat{\xi}$ có tính chất $\bar{\hat{\xi}} = 0$

$$E\hat{\xi} = 0, \operatorname{cov}(\hat{\xi}) = \sigma^2 (I_n - H) \quad (3.19)$$

c)
$$\hat{\sigma}^2 = \frac{\hat{\xi}^T \hat{\xi}}{n-k-1} = \frac{\sum_{i=1}^{n} \hat{\xi}_i^2}{n-k-1} \quad (3.20)$$

là ước lượng không chệch của σ^2, tức là $E\hat{\sigma}^2 = \sigma^2$

d) $\hat{\theta}, \hat{\xi}$ là không tương quan tức là:

$$\operatorname{cov}(\hat{\theta}, \hat{\xi}) = 0; \quad \operatorname{cov}(\hat{\theta}, \hat{\sigma}^2) = 0 \quad (3.21)$$

Chứng minh :

Tính chất a,

$$E\hat{\theta} = E\left[(X^T X)^{-1} X^T Y\right] = (X^T X)^{-1} X^T EY$$

$$= (X^T X)^{-1} X^T E(X\theta + \xi)$$

$$= (X^T X)^{-1} X^T X\theta = \theta$$

$$\operatorname{cov}(\hat{\theta}) = E\left[(X^T X)^{-1} X^T Y((X^T X)^{-1} X^T Y)^T\right]$$

$$= (X^T X)^{-1} X^T E(YY^T X(X^T X)^{-1})$$

$$= (X^T X)^{-1} X^T \operatorname{cov}(Y) X(X^T X)^{-1}$$

$$= \sigma^2 (X^T X)^{-1} X^T I_n X(X^T X)^{-1} = \sigma^2 (X^T X)^{-1}$$

Tính chất b,

Do $\hat{\xi} = (I_n - H)Y$ nên theo (3.9), (3.11), (3.13), (3.14), (3.15) có:

$$E(\hat{\xi}) = (I_n - H)EY = (I - H)X\theta = 0$$

$$\operatorname{cov}(\hat{\xi}) = E(\hat{\xi}^T \hat{\xi}) = E\left[(I_n - H)YY^T (I_n - H)^T\right]$$

$$= \sigma^2 (I_n - H)(I_n - H)^T$$

$$= \sigma^2 (I_n - H)$$

Tính chất c,

$$E(\hat{\sigma}^2) = \frac{E(\hat{\xi}^T \hat{\xi})}{n-k-1} = \frac{\sum_{i=1}^{n} E\hat{\xi}_i^2}{n-k-1}$$

Do các ξ_i là độc lập nên

$$E(\hat{\sigma}^2) = \frac{tr(\text{cov}(\hat{\xi}))}{n-k-1}$$

theo (3.19) $E\hat{\xi} = 0, \text{cov}(\hat{\xi}) = \sigma^2(I_n - H)$ nên

$$E(\hat{\sigma}^2) = \frac{tr(\text{cov}(\hat{\xi}))}{n-k-1}$$

$$= \sigma^2 \frac{tr(I_n - H)}{n-k-1} = \sigma^2 \frac{(I_n - tr(H))}{n-k-1}$$

Mặt khác theo tính chất vết của ma trận thì

$$tr(H) = tr(X(X^T X)^{-1} X^T) = tr((X^T X)^{-1} X^T X) = tr(I_{k+1}) = k+1$$

Vậy nên:

$$E(\hat{\sigma}^2) = \frac{E(\hat{\xi}^T \hat{\xi})}{(n-k-1)} = \frac{\sigma^2(n-k-1)}{(n-k-1)} = \sigma^2$$

Tính chất c được chứng minh.

Tính chất d

$$\text{cov}(\hat{\theta}, \hat{\xi}) = E\left[\left(\hat{\theta} - E\hat{\theta}\right)\left(\hat{\xi} - E\hat{\xi}\right)\right]$$

$$= E\left\{\left[(X^T X)^{-1} X^T Y - (X^T X)^{-1} X^T EY\right]\left[(I_n - H)Y - E(I_n - H)Y\right]\right\}$$

$$= E\left\{\left[(X^T X)^{-1} X^T (Y - EY)\right]\left[(I_n - H)(Y - EY)\right]\right\}$$

$$= E\left[(X^T X)^{-1} X^T (I_n - H)(Y - EY)^2\right] = 0$$

do $X^T(I_n - H)$

$$\text{cov}(\hat{\theta}, \hat{\sigma}^2) = E\left[\left(\hat{\theta} - E\hat{\theta}\right)\left(\hat{\sigma}^2 - E\hat{\sigma}^2\right)\right]$$

$$= E(\hat{\theta} - \theta)(\hat{\sigma}^2 - \sigma^2)$$

Theo các chứng minh trên thì $\hat{\theta}, \sigma^2$ lần lượt là các ước lượng không chệch của θ, σ^2 nên $\text{cov}(\hat{\theta}, \hat{\sigma}^2) = E(\hat{\theta} - \theta)(\hat{\sigma}^2 - \sigma^2) = 0$

2.2.2. Hệ số xác định R

$$R^2 = \frac{\hat{Y}^T\hat{Y} - n(\bar{y})^2}{Y^TY - n(\bar{y})^2} = \frac{\sum_{i=1}^{n}(y_i)^2 - n(\bar{y})^2}{\sum_{i=1}^{n}(y_i)^2 - n(\bar{y})^2} = \frac{s_{\hat{y}}^2}{s_y^2} \qquad (3.22)$$

gọi là bình phương của hệ số xác định, đó là tỷ lệ biến thiên của các biến Y' được giải thích bởi các biến $x_{i1}, x_{i2}, \dots, x_{ik}$ $i = 1 \dots n$.

Mà $$\sum_{i=1}^{n} y_i^2 = \sum_{i=1}^{n} \hat{y}_i^2 + \sum_{i=1}^{n} \hat{\xi}_i^2 \qquad (3.23)$$

Suy ra:

$$\sum_{i=1}^{n} \hat{\xi}_i^2 = \left[\sum_{i=1}^{n} y_i^2 - n(\bar{y})^2\right]\left[1 - R^2\right] = ns_y(1 - R^2) \qquad (3.24)$$

ta nhận được phương trình để tính sai số bình phương trung bình thông qua công thức (3.24)

Ví dụ 3.2: Theo ví dụ 3.1 ta đã tính được

$$\sum_{i=1}^{n} \hat{\xi}_i^2 = \hat{\xi}^T\hat{\xi} = Y^TY - Y^TX\hat{\theta}$$

$$= 245626 - 1696 \times 32,2777 - 2,5057 \times 35463 - 4,7587 \times 21409$$

$$= 144,3734$$

$$ns_y = \sum_{i=1}^{n} y_i^2 - n(\bar{y})^2 = 245626 - 239701 = 5924,6667$$

Thay vào công thức (3.23) ta tính được bình phương hệ số xác định

$$R^2 = \frac{ns_y - \sum_{i=1}^{n}\hat{\xi}_i^2}{ns_y} = \frac{5924,6667 - 144,3734}{5924,6667} \approx 0,9756$$

2.2.3. Ma trận tương quan

Xét mô hình hồi quy (3.3):

$$y_1 = \theta_0 + \theta_1 x_{11} + \theta_2 x_{12} + \dots + \theta_k x_{1k} + \xi_1$$

$$y_2 = \theta_0 + \theta_1 x_{21} + \theta_2 x_{22} + \dots + \theta_k x_{2k} + \xi_2 \qquad (3.3)$$

....

$$y_n = \theta_0 + \theta_1 x_{n1} + \theta_2 x_{n2} + \dots + \theta_k x_{nk} + \xi_n$$

Thoả mãn các giả thiết của mô hình hồi quy tuyến tính cổ điển.

Ký hiệu r_{pq} hệ số tương quan giữa biến X_p và X_q trong đó $p,q = 1, 2,\dots, k$.

Ở đây ta chỉ chú ý quan hệ tương quan của các biến độc lập với nhau. Hệ số tương quan giữa hai biến X_p, X_q là:

$$r_{pq}^2 = \frac{\left(\sum_{i=1}^{n}(x_{ip} - \bar{x})(x_{iq} - \bar{x})\right)^2}{\sum_{i=1}^{n}(x_{ip} - \bar{x})^2 \sum_{i=1}^{n}(x_{iq} - \bar{x})^2}$$

Ta thấy rằng $r_{pq}^2 = r_{qp}^2$ và $r_{pq}^2 = 1$.

Ta có ma trận tương quan như sau:

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1k} \\ r_{21} & r_{22} & \cdots & r_{2k} \\ \vdots & \vdots & \cdots & \vdots \\ r_{k1} & r_{k2} & \cdots & r_{kk} \end{bmatrix}$$

Ma trận tương quan R là ma trận đối xứng, tức là $R = R^T$.

2.2.4. Khoảng tin cậy cho các hệ số hồi quy θ_j

Xét mô hình tuyến tính cổ điển (3.4) với giả thiết các ξ_i , $i = 1, 2,.., n$ cùng phân bố chuẩn $N(0, \sigma^2)$ và độc lập, tức là $\xi = (\xi_1, \xi_2, \dots \xi_n)^T$ có phân bố $N_n(0, \sigma^2 I_n)$.

Mệnh đề 3.2

a) $\hat{\theta}$ có phân bố chuẩn $N_{k+1}(\theta, \sigma^2 (X^T X)^{-1})$.

b) $\frac{(n-k-1)\hat{\sigma}^2}{\sigma^2} = \sum_{i=1}^{n} \hat{\xi}_i^2$ có phân bố χ^2 với $n - k - 1$ bậc tự do.

c) $\hat{\theta}, \sigma^2$ là độc lập.

Chứng minh a

Theo (3.18) có $E\hat{\theta} = \theta$, $\text{cov}(\hat{\theta}) = \sigma^2(X^T X)^{-1}$ nên $\hat{\theta}$ có phân bố $N_{k+1}(\theta, \sigma^2(X^T X)^{-1})$.

+) Theo (3.19) có $E\hat{\xi} = 0, \text{cov}(\hat{\xi}) = \sigma^2(I_n - H)$ nên $\hat{\xi}$ có phân bố chuẩn $N_n(0, \sigma^2(I_n - H))$.

Chứng minh b.

i) Vì $(I-H)$ là ma trận lũy đẳng $(I-H)=(I-H)^2$ nên nếu ký hiệu λ, e là cặp giá trị riêng và vector riêng của $(I-H)$, ta có:

$(I-H)\,e=\lambda e \Rightarrow (I-H)^2=\lambda\,(I-H)\,e=\lambda^2 e$ vậy nên $\lambda^2 e=\lambda e$ hay $\lambda=\lambda^2$. Suy ra $\lambda=1$ hoặc $\lambda=0$.

Vì $tr(I-H)=n-k-1=\lambda_1+\ldots+\lambda_n$ nên $n-k-1$ giá trị riêng đầu tiên của $(I-H)$ là 1 còn $k+1$ giá trị riêng còn lại bằng 0.

ii) Giả sử $e_1, e_2, \ldots, e_{n-k-1}$ là $n-k-1$ vector riêng ứng với các trị riêng có giá trị là 1 còn $e_{n-k}, \ldots, e_n$ là $k+1$ trị riêng ứng với các trị riêng bằng 0 của ma trận $(I-H)$. Theo công thức khai triển phổ của ma trận ta có

$$I-H=e_1e_1^T+\ldots+e_{n-k-1}e_{n-k-1}^T$$

Đặt

$$V=\begin{bmatrix} V_1 \\ V_2 \\ \vdots \\ V_{n-k-1} \end{bmatrix}=\begin{bmatrix} e_1^T\xi \\ e_2^T\xi \\ \vdots \\ e_{n-k-1}^T\xi \end{bmatrix}$$

Khi đó V có phân bố chuẩn với $E(V)=0$ còn

$$\operatorname{cov}(V_i, V_j)=e_i^T(\sigma^2 I)e_i=\begin{cases}\sigma^2 : i=j\\ 0 : i\neq j\end{cases}$$

nên $V_1, V_2, \ldots, V_{n-k-1}$ có phân bố chuẩn độc lập $N(0, 1)$ và V có $N(0, \sigma^2 I_{n-k-1})$.

Do đó $\quad (n-k-1)\hat{\sigma}^2=\tilde{\xi}^T\tilde{\xi}=\xi^T(I-H)\xi=V^TV=V_1^2+\ldots+V_{n-k-1}^2$ và $\dfrac{(n-k-1)\hat{\sigma}^2}{\sigma^2}$ có phân bố χ^2 với $(n-k-1)$ bậc tự do. Suy ra mệnh đề đã được chứng minh.

Chứng minh c

Theo (3.21) có $\operatorname{cov}(\hat{\theta},\hat{\xi})=0$ nên $\hat{\theta},\hat{\xi}$ là độc lập.

Mệnh đề 3.3: Xét mô hình hồi quy tuyến tính cổ điển $Y=X\theta+\xi$ với X có hạng $k+1\le n$ và ξ có phân bố chuẩn $N(0,\sigma^2 I_n)$. Khi đó miền tin cậy $1-\alpha$ của θ xác định bởi :

$$(\theta-\hat{\theta})^T X^T X(\theta-\hat{\theta}) \leq (k+1)\hat{\sigma}^2 F_{k+1,n-k-1}(\alpha) \qquad (3.25)$$

Trong đó $F_{k+1,n-k-1}(\alpha)$ là phân vị trên mức α của phân bố F với $k+1$ và $n-k-1$ bậc tự do . Nói cách khác với độ tin cậy $1-\alpha$, giá trị chân thực θ phải nằm bên trong Ellipsoid

$$(x-\hat{\theta})^T X^T X(x-\hat{\theta}) \leq (k+1)\hat{\sigma}^2 F_{k+1,n-k-1}(\alpha)$$

Hơn nữa khoảng tin cậy đồng thời với mức $1-\alpha$ của các θ_j, $j = 0, 1, .., k$ được xác định bởi các mút: $\hat{\theta}_j = \sqrt{\hat{D}(\hat{\theta}_j)(k+1)F_{k=1,n-k-1}(\alpha)}$ trong đó $\hat{D}(\hat{\theta}_j)$ ký hiệu phần tử thứ j trên đường chéo chính của ma trận $\hat{\sigma}^2(X^T X)^{-1}$ và là ước lượng không chệch của $D(\hat{\theta}_j)$.

Chứng minh

Xét ma trận căn bậc hai đối xứng $(X^T X)^{1/2}$ và đặt $U=(X^T X)^{1/2}(\hat{\theta}-\theta)$

Ta có $E(U) = 0$ và

$$\text{cov}(U) = (X^T X)^{1/2}\text{cov}(\hat{\theta})(X^T X)^{1/2}$$

$$= \hat{\sigma}^2 (X^T X)^{1/2}(X^T X)^{-1}(X^T X)^{1/2} = \hat{\sigma}^2 I_{k+1}$$

Vậy U có phân bố chuẩn $N(0, \sigma^2 I_{k+1})$. Do đó:

$\frac{1}{\hat{\sigma}^2}U^T U = \frac{1}{\hat{\sigma}^2}(\hat{\theta}-\theta)^T(X^T X)(\hat{\theta}-\theta)$ có phân bố χ^2 với $k+1$ bậc tự do.

Hơn nữa theo mệnh đề 3.2 thì $\frac{(n-k-1)\hat{\sigma}^2}{\sigma^2}$ có phân bố χ^2 với $(n-k-1)$ bậc tự do và độc lập với $\hat{\theta}$, tức là độc lập với $U^T U$ vì vậy đại lượng

$$F = \frac{(\hat{\theta}-\theta)^T X^T X(\hat{\theta}-\theta)}{\hat{\sigma}^2(k+1)} = \frac{U^T U/(k+1)\hat{\sigma}^2}{(n-k-1)\hat{\sigma}^2/(n-k-1)\sigma^2}$$

có phân bố F với $k+1$ và n-k-1 bậc tự. Từ đó:

$$P\{F \leq F_{k+1,n-k-1}(\alpha)\} = 1-\alpha$$

Hoặc

$$P\left\{(\hat{\theta}-\theta)^T X^T X(\hat{\theta}-\theta) \leq (k+1)\hat{\sigma}^2 F_{k+1,n-k-1}(\alpha)\right\} = 1-\alpha$$

Nếu chiếu ellipsoid 3.25 ứng với dạng toàn phương với ma trận $X^T X$ trên các trục của nó, với xác suất $1-\alpha$ ta có:

$$P\left\{\left|\hat{\theta}_j - \theta_j\right| \le \left[(k+1)\hat{\sigma}^2 V_{jj} F_{k+1,n-k-1}(\alpha)\right]^{1/2}, j = 0..k\right\} \ge 1-\alpha$$

trong đó V_{jj} là phần tử trên đường chéo chính của $(X^T X)^{-1}$.

Nếu (3.25) xảy ra thì theo bất đẳng thức Cauchy mở rộng đặt $c^2 = (k+1)\hat{\sigma}^2 V_{jj} F_{k+1,n-k-1}(\alpha)$, $u^{(j)} = (0,0 \ldots, 0,1,0, \ldots, 0)$ là vector đơn vị với 1 ở vị trí thứ j

$$\left[(\hat{\theta}-\theta)^T u^{(j)}\right]^2 \le (\hat{\theta}-\theta)^T (X^T X)(\hat{\theta}-\theta) u^{(j)T} (X^T X)^{-1} u^{(j)} \le c^2 u^{(j)T} (X^T X)^{-1} u^{(j)}$$

Hoặc $(\hat{\theta}_j - \theta_j)^2 \le c^2 V_{jj}$ với mọi $j = 0, 1, .., k$.

Mệnh đề 3.4.

Giả sử $t_{n-k-1}\left(\frac{\alpha}{2(k+1)}\right)$ là phân vị trên mức $\frac{\alpha}{2(k+1)}$ của phân bố Student với n-k-1 bậc tự do. Khi đó đồng thời ta có khoảng tin cậy của θ_i với mức tin cậy $1-\alpha$ cho bởi các đầu mút

$$\hat{\theta}_j \pm t_{n-k-1}\left(\frac{\alpha}{2(k+1)}\right)\sqrt{\hat{D}(\hat{\theta}_j)} \tag{3.26}$$

Vì $(\hat{\theta}-\theta)$ có phân bố chuẩn $N(0,\sigma^2(X^T X)^{-1})$ nên $(\hat{\theta}_j - \theta_j)$ có phân bố chuẩn $N(0, \sigma^2 V_{jj})$ và độc lập với $\hat{\sigma}^2$, trong đó $\frac{(n-k-1)\hat{\sigma}^2}{\sigma^2}$ có phân bố χ^2 với $n-k-1$ bậc tự do. Vì vậy biến

$$T_j = \frac{(\hat{\theta}_j - \theta_j)/V_{jj}^{1/2}}{\left\{(n-k-1)\hat{\sigma}^2 /(n-k-1)\sigma^2\right\}^{1/2}} = \frac{(\hat{\theta}_j - \theta_j)}{\sigma^2 V_{jj}^{1/2}} = \frac{(\hat{\theta}_j - \theta_j)}{\sqrt{\hat{D}(\hat{\theta}_j)}}$$

có phân bố Student với $(n-k-1)$ bậc tự do, vậy nên:

$$P(A_j) := P\left\{\left|T_j\right| < t_{n-k-1}\left(\frac{\alpha}{2(k+1)}\right)\right\} = 1 - \frac{\alpha}{(k+1)}; j = 0,1,\ldots,k$$

trong đó $A_j = \left|T_j\right| < t_{n-k-1}\left(\frac{\alpha}{2(k+1)}\right)$ là biến cố ở trong xác suất thứ hai.

Hơn nữa

$$
\begin{aligned}
P(A_0A_1A_2...A_k) &= 1 - P(\overline{A_0A_1A_2...A_k}) \\
&= 1 - P(\overline{A_0} \cup \overline{A_1} \cup \overline{A_2} \cup ... \cup \overline{A_k}) \\
&\geq 1 - P(\overline{A_0}) - P(\overline{A_1}) - P(\overline{A_2}) - ... - P(\overline{A_k}) \\
&= 1 - \underbrace{\left(\frac{\alpha}{(k+1)} + \frac{\alpha}{(k+1)} + ... + \frac{\alpha}{(k+1)}\right)}_{k+1} \\
&= 1 - \alpha
\end{aligned}
$$

Từ đó ta suy ra:

$$P\left\{\left|\hat{\theta}_j - \theta_j\right| \leq \hat{\sigma}^2 V_{jj} t_{n-k-1}\left(\frac{\alpha}{2(k+1)}\right); j = 0, 1, ..., k\right\} = 1 - \alpha$$

2.2.5. Kiểm định giả thiết về các hệ số hồi quy

Trong khi thiết lập mô hình hồi quy tuyến tính cổ điển (3.2), ta giả thiết tất cả các biến độc lập đều tham gia vào phương trình hồi quy. Tuy nhiên không phải các biến đều tham gia vào mô hình (3.2), tức là có một số hệ số của chúng bằng 0 nhưng giá trị ước lượng lại khác khác 0. Vậy nên cần kiểm định.

Kiểm định $H_0 : \theta_{p-1} = ... = \theta_k = 0 \ (0 \leq p \leq k)$ (3.27)

có nghĩa là các biến độc lập $X_{p+1}, ..., X_k$.

Với đối thuyết $H_1 : \exists_i \in \{p + 1, ..., k\}$ sao cho $\theta_i \neq 0$, tức là có ít nhất một trong các biến này cần tính đến trong mô hình.

Kiểm định H_0 dạng

$$H_0 : \begin{cases} c_{10} + c_{11}\theta_1 + ... + c_{1k}\theta_k = a_1 \\ c_{20} + c_{21}\theta_1 + ... + c_{2k}\theta_k = a_2 \\ \dots\dots\dots \\ c_{k-p,0} + c_{k-p,1}\theta_1 + ... + c_{k-p\,k}\theta_k = a_{k-p} \end{cases} \quad (3.28)$$

$$\Leftrightarrow C\theta = a;\ C = \left[c_{ij}\right]_{(k-p)\times(k+1)};\quad a = (a_1, a_2, ..., a_{k-p})^T$$

Giả thiết H_0 xác định bởi (3.27) là trường hợp riêng của (3.28) khi ma trận C có dạng:

$$C = \begin{bmatrix} 0 & 0 & \dots & 0 & \vdots & 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & \vdots & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & 0 & \vdots & 0 & 0 & \dots & 1 \end{bmatrix} = \left[0 \vdots I_{k-p}\right]; a = [0,0,..,0]^T$$

Theo mệnh đề 3.2 thì $\hat{\theta}$ có phân bố chuẩn $N_{k+1}(\theta, \sigma^2(X^TX)^{-1})$ nên $C\hat{\theta}$ chính là ước lượng không chệch của $C\theta$ và có phân bố chuẩn $N_{k-p}(C\theta, \sigma^2 C(X^TX)^{-1}C^T)$ nên ta sẽ bác bỏ H_0 : $C\theta = \alpha$ nếu a nằm ngoài ellipsoid tin cậy của $C\theta$.

Quy tắc kiểm định: bác bỏ $H_0: C\theta = \alpha$ nếu

$$\frac{(C\hat{\theta})(C(X^TX)^{-1}C^T)^{-1}C\hat{\theta}}{\hat{\sigma}^2} > (k-p)F_{k-p,n-k-1}(\alpha) \tag{3.29}$$

2.3. Kiểm tra sự phù hợp của mô hình

Mô hình hồi quy tuyến tính (3.3) sẽ phù hợp với dãy số liệu đang quan sát nếu các biến nhiễu ξ_i chỉ do các yếu tố ngẫu nhiên gây ra, chúng là các biến ngẫu nhiên độc lập cùng phân phối. Do mô hình đang xét có phân phối chuẩn nên ta cần kiểm tra các ξ_i có phân phối chuẩn $\xi \sim N(0, \sigma^2)$ hay không? Để kiểm tra các giả thiết đó thường xét các tiêu chuẩn sau:

2.3.1. Tiêu chuẩn F

Xét đại lượng: $F = \dfrac{(n-k-1)R^2}{k(1-R^2)}$ (3.30)

Nếu sai số ξ có phân bố $N_n(0, \sigma^2 I_n)$ và nếu $\theta_j = 0; j = 0, 1, \dots, k$ thì F đã cho bởi (3.29) có phân bố F với k và n - k - 1 bậc tự do.

Chứng minh:

Ta có

$$\sum_{i=1}^{n}\left(\xi_i-\bar{\xi}\right)^2=\sum_{i=1}^{n}\xi_i^2-2\bar{\xi}\sum_{i=1}^{n}\xi_i+n(\bar{\xi})^2$$

$$=\sum_{i=1}^{n}\xi_i^2-2n(\bar{\xi})^2+n(\bar{\xi})^2$$

$$=\sum_{i=1}^{n}\xi_i^2-n(\bar{\xi})^2=\sum_{i=1}^{n}y_i^2-n(\bar{y})^2$$

Do $\theta_j = 0$; $j = 0, 1, \ldots, k$. Mà $Y^TY-n(\bar{y})^2=\sum_{i=1}^{n}y_i^2-n(\bar{y})^2$ nên

$$\frac{Y^TY-n(\bar{y})^2}{\sigma^2}=\frac{\sum_{i=1}^{n}\left(\xi_i-\bar{\xi}\right)^2}{\sigma^2}$$ có phân phối χ^2_{n-1}.

$$\Rightarrow\frac{Y^TY-n(\bar{y})^2-\hat{\xi}^T\hat{\xi}}{\sigma^2}=\frac{ns_y^2-ns_y^2(1-R^2)}{\sigma^2}=ns_y^2R^2$$ có phân phối χ^2 với

$n-1-n+k+1=k$ bậc tự do và độc lập với $\hat{\xi}^T\hat{\xi}$.

$$F=\frac{ns_y^2R^2/k}{ns_y^2(1-R^2)/(n-k-1)}=\frac{R^2(n-k-1)}{k(1-R^2)}$$ có phân bố F với k và $n-k-1$

bậc tự do

Từ mệnh đề này ta đưa ra quy tắc để kiểm định: nếu F quá lớn hoặc F gần 0 ta cần bác bỏ giả thiết ξ có phân bố chuẩn hoặc các hệ số $\theta_j = 0$; $j = 0, 1, \ldots, k$

Ví dụ 3.3: Xét ví dụ 3.1; 3.2 với n = 12; k = 2. Ta có

$$F=\frac{(n-k-1)R^2}{k(1-R^2)}=\frac{(12-2-1)0{,}9756}{2(1-0{,}9756)}\approx 179{,}9621$$

Tra bảng phân vị Fisher với mức ý nghĩa α =0,05

$F_{(2,9;\,0,95)} = 4{,}26 \Rightarrow F > F_{(2,9;\,0,95)} = 4{,}26$. Vậy nên bác bỏ H_0: các hệ số hồi quy bằng 0.

2.3.2. Kiểm tra $D\xi = \sigma^2$ nhưng σ^2 chưa biết

Ta đưa ra giả thuyết H_0: $D\xi = \sigma^2$

Với đối thuyết H_1: $D\xi \neq \sigma^2$

Tại mỗi điểm thí nghiệm thứ i ($i = 1, 2, .., n$) ta lặp lại m lần đo được m giá trị của $y_{i1}, y_{i2}, \ldots, y_{im}$.. Ta coi đó là một mẫu và tính được

$$\overline{y}_i = \frac{1}{m}\sum_{t=1}^{m} y_{it}; \quad s_i^2 = \frac{1}{m-1}\sum_{t=1}^{m}(y_{it} - \overline{y}_i)^2; \quad i = \overline{1,n}$$

Các s_i^2 đều là ước lượng trúng của $D\xi = \sigma^2$. Vậy nên nếu H_0 đúng thì n số s_i^2 phải bằng nhau. Để kiểm định H_0 ta sẽ kiểm định sự bằng nhau của n phương sai: $s_1^2 = s_2^2 = \ldots = s_n^2$

Dùng thống kê:

$$G = \frac{s_{max}^2}{s_1^2 + \ldots + s_n^2} \text{ trong đó: } s_{max}^2 = \max_i s_i^2$$

G ~ Cochran;

Chọn α, tra bậc tử = $m-1$; tra bậc mẫu = n ta được G_α. Tính $\hat{G}$

+) Nếu $\hat{G} < G_\alpha$ => Công nhận H_0 đúng. Dùng phương sai tái sinh $s_{ts}^2 = \frac{1}{n}\sum_{i=1}^{n} s_i^2$ để ước lượng σ^2.

+) Nếu $\hat{G} \geq G_\alpha$ thì ta bác bỏ H_0.

2.3.3. Kiểm định sự phù hợp của mô hình vừa tìm được

*) Xem số biến có thực sự bằng k hay không? những biến nào ảnh hưởng trực tiếp đến $\hat{Y}$.

$$Y = \theta_0 + \theta_1 X_1 + \ldots + \theta_k X_k \text{ xem } \theta_j = 0 \ (j = 1, 2, \ldots, k)$$

Giả thiết H_0: $\theta_j = 0$ ($j = 1, 2, \ldots, k$).

Đối thuyết H_1: $\theta_j \neq 0$.

Chọn thống kê

$$t_{\hat{\theta}_j} = \frac{\hat{\theta}_j - \theta_j}{\sqrt{\hat{D}(\hat{\theta}_j)}} = \frac{\hat{\theta}_j}{\sqrt{\hat{D}(\hat{\theta}_j)}} (\text{do } \theta_j = 0)$$

trong đó: $\hat{D}(\hat{\theta}_j) = \hat{\sigma}^2 \left\{(X^T X)^{-1}\right\}_{jj}$

$$\hat{\sigma}^2 = \frac{\hat{\xi}^T \hat{\xi}}{n-k-1} = \frac{\sum_{i=1}^{n} \hat{\xi}_i^2}{n-k-1}$$

Theo mệnh đề 3.4 $t_{\hat{\theta}_j}$ tuân theo phân bố Student với bậc $n - k - 1$.

Chọn mức ý nghĩa. Tra bảng ta sẽ được t_α

+) Nếu $\left|t_{\hat{\theta}_j}\right| < t_\alpha$ công nhận $H_0 \Rightarrow \theta_j = 0$

+)Nếu $\left|t_{\hat{\theta}_j}\right| \geq t_\alpha$ thì bác bỏ $H_0 \Rightarrow \theta_j \neq 0$

Nếu $\theta_j = 0$ thì các hệ số còn lại phải tính lại vì các hệ số có thể có tương quan với nhau

*) Sự phù hợp của $\hat{Y}$

Ta có $s_{du}^2 = \frac{1}{n-k-1} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$ cũng là ước lượng của σ^2 nhưng phụ thuộc vào dạng của $\hat{Y}$.

Nếu $\hat{Y}$ phù hợp với hình nghiên cứu thì hai phương sai bằng nhau.

Kiểm định giả thuyết H_0: $\hat{\sigma}^2 = s_{du}^2$

đối thuyết H_1: $\hat{\sigma}^2 \neq s_{du}^2$

Chọn thống kê $F = \frac{s_{du}^2}{\hat{\sigma}^2} (\hat{\sigma}^2 \leq s_{du}^2)$ ~ Fisher bậc tử là $n - k - 1$; bậc mẫu là $n(m-1)$.

Chọn α tra bảng tìm F_α. Quy tắc kiểm định như sau:

+) Nếu $F < F_\alpha \Rightarrow$ công nhận $H_0 \Rightarrow$ mô hình phù hợp

+) Nếu $F \geq F_\alpha \Rightarrow$ Bác bỏ $H_0 \Rightarrow$ mô hình không phù hợp, phải giả thiết lại.

Ví dụ 3.4: Xét tiếp ví dụ 3.1

*) Kiểm định các hệ số θ_j

Theo tính chất của ước lượng bình phương cực tiểu ta có

$$\hat{\sigma}^2 = \frac{\hat{\xi}^T \hat{\xi}}{n-k-1} = \frac{\sum_{i=1}^{n} \hat{\xi}_i^2}{n-k-1} = \frac{144,3734}{9} = 16,0415$$

$\Rightarrow$ ta sẽ tính được ước lượng cho phương sai của các hệ số là:

$$\hat{D}(\hat{\theta}_0) = \hat{\sigma}^2.\left\{(X^T X)^{-1}\right\}_{00} = 16,04515 \times 2,439963 = 39,1407$$

$$\Rightarrow \sqrt{\hat{D}(\hat{\theta}_0)} = 6,2562$$

$$\hat{D}(\hat{\theta}_1) = \hat{\sigma}^2.\left\{(X^T X)^{-1}\right\}_{11} = 16,04515 \times 0,006757 = 0,1083$$

$$\Rightarrow \sqrt{\hat{D}(\hat{\theta}_1)} = 0,3287$$

$$\hat{D}(\hat{\theta}_2) = \hat{\sigma}^2.\left\{(X^T X)^{-1}\right\}_{22} = 16,04515 \times 0,010509 = 0,1686$$

$$\Rightarrow \sqrt{\hat{D}(\hat{\theta}_2)} = 0,4106$$

Tính thống kê :

$$t_{\hat{\theta}_j} = \frac{\hat{\theta}_j}{\sqrt{\hat{D}(\hat{\theta}_j)}}$$

$$t_{\hat{\theta}_0} = \frac{\hat{\theta}_0}{\sqrt{\hat{D}(\hat{\theta}_0)}} = \frac{32,2777}{6,2526} = 5,1623$$

$$t_{\hat{\theta}_1} = \frac{\hat{\theta}_1}{\sqrt{\hat{D}(\hat{\theta}_1)}} = \frac{2,5057}{0,3287} = 7,6231$$

$$t_{\hat{\theta}_2} = \frac{\hat{\theta}_2}{\sqrt{\hat{D}(\hat{\theta}_2)}} = \frac{4,7587}{0,4106} = 11,5896$$

Chọn α = 0,05, tra bảng Student với $t_{n-k-1}\left(\frac{\alpha}{2(k+1)}\right) = t_9\left(\frac{0,05}{2.3}\right) \approx 2,933$

=> $t_{\hat{\theta}_j} > t_{n-k-1}\left(\frac{\alpha}{2(k+1)}\right)$ vậy nên bác hỏ H_0 là các $\theta_j = 0$.

*) Kiểm định sự phù hợp của $\hat{Y}$

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} = 141,3332$$

$$=> s_Y^2 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n-1} = \frac{5924,6667}{11} = 538,6061$$

$$=> \hat{F} = \frac{s_Y^2}{\hat{\sigma}^2} = \frac{538,6061}{16,0415} = 35,5758$$

Chọn $\alpha = 0,05$, với bậc tử là $n - 1 = 11$; bậc mẫu là $n - k - 1 = 9$. Tra bảng Fisher có $F_{(11,\ 9,\ 0,05)} \approx 3,1 \Rightarrow \hat{F} > F_{(11,\ 9;\ 0,05)}$. Vậy nên bác bỏ H_0. Mô hình cần phải giả thiết lại.

§ 3. TUYẾN TÍNH HÓA MỘT SỐ HÀM PHI TUYẾN

3.1. Hàm đa thức một biến bậc *n*

$$\tilde{y} = a_0 + a_1 x + a_2 x^2 ... + a_n x^n$$

Bằng cách đặt: $X_j = x^j,\ j = \overline{1,n}$ thu được

$\tilde{y} = a_0 + a_1 X_1 + a_2 X_2 ... + a_n X_n$ là mô hình tuyến tính *n* biến.

3.2. Hồi quy parabol bội

$$\tilde{y} = a_0 + \frac{a_1}{x_1} + \frac{a_2}{x_2} + ... + \frac{a_n}{x_n}.$$

Bằng cách đặt $X_j = \frac{1}{x_j};\ j = \overline{1,n}$ thu được

$\tilde{y} = a_0 + a_1 X_1 + a_2 X_2 ... + a_n X_n$ là mô hình tuyến tính *n* biến.

3.3. Hồi quy toàn phương

$$\tilde{y} = a_0 + \sum_{j=1}^{n} a_j x_j + \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j$$

Tuyến tính hoá bằng cách đặt $X_j = x_j;\ j = \overline{1,n}$ và $X_{ij} = x_i x_j;\ i = \overline{1,n}; j = \overline{1,n}$

3.4. Hồi quy logarit

$$\tilde{y} = a_1 + a_1 \ln x_1 + a_2 \ln x_2 ... + a_n \ln x_n$$

Bằng cách đặt $X_j = \ln x_j,\ j = \overline{1,n}$ ta thu được

$\tilde{y} = a_0 + a_1X_1 + a_2X_2 ... + a_nX_n$ là mô hình tuyến tính n biến

3.5. Hồi quy căn

$$\tilde{y} = a_0 + a_1x_1^{1/2} + a_2x_2^{1/2} ... + a_nx_n^{1/2}$$

Bằng cách đặt $X_j = x_j^{1/2}$, $j = \overline{1,n}$ thu được

$\tilde{y} = a_0 + a_1X_1 + a_2X_2 ... + a_nX_n$ là mô hình tuyến tính n biến.

3.6. Hồi quy luỹ thừa

$$\tilde{y} = a_0x_1^{a1}x_2^{a2}...x_n^{an}$$

Bằng cách lấy logarit hai vế

$$\ln\tilde{y} = \ln a_0 + a_1\ln x_1 + a_2\ln x_2 ... + a_n\ln x_n$$

Đặt $Y = \ln\tilde{y}$, $X_j = \ln x_j$, $j = \overline{1,n}$ thu được

$$\tilde{y} = \ln a_0 + a_1X_1 + a_2X_2 + ... + a_nX_n$$

là mô hình tuyến tính n biến.

3.7. Hồi quy mũ

$$\tilde{y} = e^{a_0 + a_1x + a_2x^2 ... + a_nx^n}$$

Lấy logarit hai vế được mô hình tuyến tính hai biến

$$Y^* = \ln\tilde{y} = a_0 + a_1x_1 + a_2x_2 ... + a_nx_n$$

3.8. Hồi quy nghịch đảo

$$\tilde{y} = \frac{1}{a_0 + a_1x + a_2x^2 ... + a_nx^n}$$

Tuyến tính hoá bằng cách đặt: $Y^* = \dfrac{1}{\tilde{y}}$ và $X_y = x^y$, $y = \overline{1,n}$

3.9. Hồi quy mũ nghịch đảo

$$\tilde{y} = \frac{1}{1 + e^{a_0 + a_1x + a_2x^2 ... + a_nx^n}}$$

Tuyến tính hoá bằng cách đặt

$$Y^* = \ln(\frac{1}{\tilde{y}} - 1) = a_0 + a_1x_1 + a_2x_2 ... + a_nx_n$$

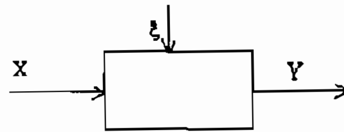
§4. ÁP DỤNG PHƯƠNG PHÁP BPCT CHO HÀM MỘT BIẾN

4.1. Hàm tuyến tính

Giả thiết

$$\begin{aligned} &Y = \theta_0 + \theta_1 X + \xi \\ &\xi \sim N(0,1) \end{aligned} \tag{3.31}$$

Tiến hành n thí nghiệm ta được bảng 3.3, hình 3.1



Hình 3.1.

Bảng 3.3

N^0	X_0	X	Y
1	1	x_1	y_1
2	1	x_2	y_2
...	...	...	...
n	1	x_n	y_n

*) Ước lượng các hệ số θ

Áp dụng phương pháp bình phương cực tiểu để ước lượng các hệ số θ.

Để $S = \sum_{i=1}^{n}(y_i - \hat{\theta}_0 - \theta_1 x_i)^2 \rightarrow \min$ thì

$$\begin{cases} \dfrac{\partial S}{\partial \theta_0} = 0 \\ \dfrac{\partial S}{\partial \theta_1} = 0 \end{cases} \Leftrightarrow \begin{cases} 2\sum_{i=1}^{n}(y_i - \theta_0 - \theta_1 x_i)(-1) = 0 \\ 2\sum_{i=1}^{n}(y_i - \theta_0 - \theta_1 x_i)(-x_i) = 0 \end{cases} \Leftrightarrow \begin{cases} n\theta_0 + \theta_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n}\theta_0 x_i + \theta_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i x_i \end{cases}$$

(3.32)

Theo phương pháp BPCT thì

$$X^T X = \begin{bmatrix} \sum_{i=1}^{n} x_0 & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{bmatrix}; \ (X^T X)^{-1} = \frac{1}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2} \begin{bmatrix} \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i \\ -\sum_{i=1}^{n} x_i & n \end{bmatrix}$$

$$X^T Y = \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} y_i x_i \end{bmatrix} \quad \hat{\theta} = (X^T X)^{-1}(X^T Y)$$

$$\Rightarrow \begin{bmatrix} \hat{\theta}_0 \\ \hat{\theta}_1 \end{bmatrix} = \frac{1}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2} \begin{bmatrix} \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i \\ -\sum_{i=1}^{n} x_i & n \end{bmatrix} \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} y_i x_i \end{bmatrix}$$

Vậy nên ước lượng cho các hệ số của mô hình (3.32) là:

$$\hat{\theta}_0 = \frac{\sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i x_i}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}; \quad \hat{\theta}_1 = \frac{n\sum_{i=1}^{n} y_i x_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}$$

Để dễ tính ta cần lập bảng 3.4.

Bảng 3.4

N^0	x_i	y_i	x_i^2	$x_i y_i$
1	X_1	y_1	x_1^2	$x_1 y_1$
2	x_2	y_2	x_2^2	$x_2 y_2$
:	:	:		:
n	x_n	y_n	x_n^2	$x_n y_n$
Σ				

*) Kiểm định: giống như trường hợp *k* biến.

Ở đây cần lưu ý:

$$\hat{D}(\hat{\theta}_0) = s_{ts}^2 \frac{\sum_{i=1}^{n} x_i^2}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}$$

$$\hat{D}(\hat{\theta}_1) = s_{ts}^2 \frac{n}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}$$

Khi tính s_{du}^2 ta có:

$$s_Y^2 \le s_{du}^2 = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

Ví dụ 3.5: Tìm phương trình về sự phù thuộc của độ bền của sợi vào độ ẩm không khí (KK). Bảng 3.5 là bảng số liệu.

Bảng 3.5

Độ ẩm KK	Độ bền của sợi y						Độ bền TB
36	2	2,1	2,1	2	2	1,9	2
38	2,4	2,5	2,5	2,4	2,4	2,6	2,5
40	2,2	2,2	2,2	2,3	2,3	2,3	2,3
58	2,8	3,0	3,0	2,7	2,7	2,8	2,8
70	3,0	3,1	3,1	3,1	3,1	3,1	3,1
80	3,1	3,1	3,1	3,1	3,1	2,9	3,1
82	3,2	3,2	3,2	3,2	3,2	3,2	3,2
93	2,9	3,0	3,0	3,1	3,0	3,1	3,0

Giả sử phụ thuộc giữa độ bền của sợi với độ ẩm của KK theo mô hình sau:

$$Y = \theta_0 + \theta_1 X + \xi$$

*) Ước lượng các hệ số của mô hình:

Từ số liệu mà đầu bài cho, ta lập bảng 3.6.

Bảng 3.6

x_i	y_i	x_i^2	$x_i y_i$
36	2	1296	72,0
38	2,5	1444	95,0
40	2,3	1600	92,0
58	2,8	3364	162,4
70	3,1	4900	217,0
80	3,1	6400	248,0
82	3,2	6724	262,4
93	3,0	8649	279,0
497	22,0	34377	1427,8

*) Ước lượng các hệ số:

$s^2_1 = 0{,}008$ $s^2_2 = 0{,}008$ $s^2_3 = 0{,}008$ $s^2_4 = 0{,}02$ $s^2_5 = 0{,}004$

$s^2_6 = 0{,}016$ $s^2_7 = 0{,}0$ $s^2_8 = 0{,}008$

$$\hat{\theta}_0 = \frac{\sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i x_i}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2} = \frac{34377.22{,}0 - 497.1427{,}8}{8.34377 - 497^2} \approx 1{,}667$$

$$\hat{\theta}_1 = \frac{n\sum_{i=1}^{n} y_i x_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2} = \frac{8.1427{,}8 - 497.22{,}0}{8.34377 - 497^2} \approx 0{,}0174$$

Vậy phương trình hồi quy thực nghiệm có dạng $\hat{Y} = 1{,}667 + 0{,}0174X$

*) Kiểm định

+) Kiểm tra $D\xi = \sigma^2$

Theo bảng ta thấy có $m = 6$ thí nghiệm tại mỗi điểm i ($i = 1, 2, \ldots, 8$) và đã có hết các giá trị trung bình của Y tại mỗi điểm.

Ta tính : $s_i^2 = \dfrac{1}{m-1}\sum_{i=1}^{n}(y_i - \bar{y}_i)$ $\quad i = \overline{1,n}$

Tính các s_i^2

$$s^2_{max} = 0,02$$

$$=> \hat{G} = \frac{s^2_{\max}}{\sum_{i=1}^{n} s_i^2} \approx 0,2778$$

Chọn α = 0,05; bậc của tử = m – 1 = 5; bậc mẫu n = 8. Tra bảng Cochran G_α = 0,3595. Ta có $\hat{G} < G_\alpha$ vậy $D\xi = \sigma^2$

Dùng phương sai tái sinh :

$$s^2_{ts} = \frac{1}{n}\sum_{i=1}^{n} s_i^2 = \frac{0,072}{8} = 0,009 \text{ để ước lượng } \sigma^2$$

Kiểm tra $\theta_j = 0$: $t_{\hat{\theta}_j} = \frac{\hat{\theta}}{\sqrt{\hat{D}(\hat{\theta}_j)}}$

$$\hat{D}(\hat{\theta}_0) = s^2_{ts} \frac{\sum_{i=1}^{n} x_i^2}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2} = 0,009 \frac{34377}{8.34377 - 497^2} = 0,011$$

$$=> \sqrt{\hat{D}(\hat{\theta}_0)} = 0,105$$

$$\hat{D}(\hat{\theta}_1) = s^2_{ts} \frac{n}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2} = 0,009 \frac{8}{8.34377 - 497^2} = 0,000003$$

$$=> \sqrt{\hat{D}(\hat{\theta}_1)} = 0,00158$$

$$=> t_{\hat{\theta}_0} = \frac{\hat{\theta}_0}{\sqrt{\hat{D}(\hat{\theta}_0)}} = \frac{1,667}{0,105} = 5,87; \quad t_{\hat{\theta}_1} = \frac{\hat{\theta}_1}{\sqrt{\hat{D}(\hat{\theta}_1)}} = \frac{0,0174}{0,00158} = 11,012$$

Chọn α = 0,05, vậy bậc tự do là bậc của phương sai tái sinh = n.(m – 1) = 8.(6 – 5) = 50 ⇒ t_α = 1,68.

Ta có $\begin{cases} t_{\hat{\theta}_0} \geq t_\alpha => \theta_0 \neq 0; \\ t_{\hat{\theta}_0} > t_\alpha => \theta_1 \neq 0 \end{cases}$. Vậy phương trình hồi quy có cả θ_0; θ_1

+) Kiểm tra sự phù hợp của mô hình:

Ta phải tính các phương sai dư. Để tiện tính toán ta lập bảng 3.7.

Bảng 3.7

x_i	y_i	x_i^2	$x_i y_i$	$\hat{y}_i$
36	2	1296	72,0	2.318
38	2,5	1444	95,0	2.354
40	2,3	1600	92,0	2.390
58	2,8	3364	162,4	2.214
70	3,1	4900	217,0	2.930
80	3,1	6400	248,0	3.110
82	3,2	6724	262,4	3.146
93	3,0	8649	279,0	3.344
Σ 497	Σ 22,0	Σ 34377	Σ 1427,8	

$$s_{du}^2 = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \frac{0,288}{6} = 0,048$$

Tính thống kê

$$\hat{F} = \frac{s_{du}^2}{s_{ts}^2} = \frac{0,048}{0,009} = 5,33$$

Chọn $\alpha = 0,05$; bậc tử là $n - 2 = 6$; bậc mẫu là $n(m - 1) = 40$ => $F_\alpha = 2,34$ => $\hat{F} > F_\alpha$. Vậy nên mô hình tuyến tính mà ta giả sử là không phù hợp.

4.2. Tuyến tính hoá một số hàm phi tuyến

4.2.1. Hàm số mũ $y = ab^x + \xi$

$$\tilde{y} = ab^x \tag{3.33}$$

Lấy lg hai vế của (3.33) ta có:

$\lg\tilde{y} = \lg a + x\lg b$. Đặt $Y = \lg\tilde{y}$; $\theta_0 = \lg a$; $\theta_1 = \lg b$; $X = x$

Lúc đó (3.33) trở thành $Y = \theta_0 + \theta_1 X$ (3.34)

Sau đó áp dụng phần 4.1 để tính.

Chú ý rằng ma trận X, Y sẽ lần lượt như sau:

$$X=\begin{bmatrix}1 & x_1\\ 1 & x_2\\ \vdots & \vdots\\ 1 & x_n\end{bmatrix};Y=\begin{bmatrix}\lg y_1\\ \lg y_2\\ \vdots\\ \lg y_n\end{bmatrix}$$. Ta sẽ tính được $\hat{\theta}=(X^TX)^{-1}(X^TY)$

Cũng có thể giải trực tiếp hệ phương trình:

$$\begin{cases}\lg b\sum_{i=1}^{n}x_i^2+\lg a\sum_{i=1}^{n}x_i=\sum_{i=1}^{n}x_i\lg y_i\\ \lg b\sum_{i=1}^{n}x_i+n\lg a=\sum_{i=1}^{n}\lg y_i\end{cases}\tag{3.35}$$

Giải hệ (3.35) ta được các hệ số a,b

4.2.2. *Hàm* $\tilde{y}=\theta_0+\theta_1x^2$

Đặt $\tilde{Y}=\tilde{y}; X=x^2$ ta được $\tilde{Y}=\theta_0+\theta_1X$.

Áp dụng phần 4.1 để tính ta được θ_0; θ_1

Chú ý rằng ma trận X, Y sẽ như sau:

$$X=\begin{bmatrix}1 & x_1^2\\ 1 & x_2^2\\ \vdots & \vdots\\ 1 & x_n^2\end{bmatrix};Y=\begin{bmatrix}y_1\\ y_2\\ \vdots\\ y_n\end{bmatrix}$$

Hoặc cũng có thể giải trực tiếp từ hệ:

$$\begin{cases}\hat{\theta}_1\sum_{i=1}^{n}x_i^4+\hat{\theta}_0\sum_{i=1}^{n}x_i^2=\sum_{i=1}^{n}x_i^2y_i\\ \hat{\theta}_1\sum_{i=1}^{n}x_i^2+n\hat{\theta}_0=\sum_{i=1}^{n}y_i\end{cases}\tag{3.36}$$

Giải hệ ta sẽ tìm được các ước lượng cho $\hat{\theta}_0;\hat{\theta}_1$

4.3. Tổng quát của mô hình một biến

Trong trường hợp đã biết dạng của $f(x)$ thì đưa về các trường hợp đã làm ở trên.

Trường hợp không biết dạng của $f(x)$ thì với một số giả thiết nào đó của $f(x)$ ta xấp xỉ bằng khai triển Taylo lấy $k + 1$ số hạng

$$f(x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \ldots + \theta_k x^k$$

giả thiết $f(x)$ có đạo hàm đến cấp $k + 1$.

Vì vậy ta có thể giả sử

i) $y = \beta_0 + \beta_1 x + \theta_2 x^2 + \ldots + \beta_k x^k + \xi$ (k và β_i $i = 1,2, \ldots, k$ là chưa biết)

ii) $\xi \sim N(0, \sigma^2)$

Cần giải quyết hai vấn đề: Xác định β_i và k để đa thức thu được phù hợp với đối tượng nghiên cứu, số lượng thực nghiệm đã thu được.

4.3.1. Xác định β_i

Đặt $x_i = x^i$, $i = 1, 2, .., k$

Lúc này $y = \beta_0 + \beta_1 x_1 + \theta_2 x_2 + \ldots + \beta_k x_k + \xi$ tuyến tính với k biến. Ta sẽ sử dụng mô hình hồi quy tổng quát để tính các hệ số β_i.

Kiểm định các giả thiết xem $\beta_i = 0$ hay không?

Kiểm định xem $\hat{y}$ phù hợp với số liệu thực nghiệm hay không?

4.3.2. Xác định bậc k

Có hai cách để xác định bậc k.

*) Phương pháp 1: Kiểm định

$$\beta_k = 0 => \hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x + \hat{\beta}_2 x^2 + \ldots + \hat{\beta}_{k-1} x^{k-1}$$

*) Phương pháp 2: Dựa vào so sánh hai phương sai. Giả thiết bậc k_0 nào đó đã biết.

(1) $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x + \hat{\beta}_2 x^2 + \ldots + \hat{\beta}_{k_0} x^{k_0}$

(2) $\hat{y}' = \hat{\beta}_0 + \hat{\beta}_1 x + \hat{\beta}_1 x^2 + \ldots + \hat{\beta}_{k_0+1} x^{k_0+1}$

So sánh (1),(2) xem xét kết quả nào tốt hơn dựa vào hai phương sai

$$s_{k_0}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_1 - \hat{y}_i)^2; \quad s_{k_0+1}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_1 - \hat{y}_i')^2$$

về lý thuyết phương sai càng bé càng tốt.

Kiểm định :

+) Giả thuyết H_0: $s_{k_0}^2 = s_{k_0+1}^2$ thì mô hình (1) và (2) bằng nhau => Tăng bậc k cũng không lợi gì.

+) Đối thuyết H_1: $s_{k_0}^2 \neq s_{k_0+1}^2$ hai mô hình không bằng nhau => Tăng bậc k sẽ có lợi.

Giải thích: khi tăng bậc lên mô hình đã phát triển các hệ số β_i cũng tăng theo => để tính β_i phải tính lại từ đầu dựa trên các phương trình cơ bản, ma trận X phải thêm vào một cột.

CHƯƠNG 4

MỘT SỐ TRƯỜNG HỢP MỞ RỘNG CỦA HỒI QUY BỘI

§1. HỒI QUY VỚI BIẾN GIẢ

Trong các mô hình hồi quy tuyến tính mà chúng ta đã xem xét từ các chương trước cho đến nay thì các biến giải thích đều là các biến số lượng. Các biến đó có thể nhận giá trị bằng số. Nhưng trên thực tế có nhiều trường hợp các biến giải thích thậm chí ngay cả biến phụ thuộc là biến chất lượng. Chính vì thế mà trong phần này ta sẽ nghiên cứu khi biến giải thích là biến chất lượng.

1.1. Bản chất của biến giả - Mô hình trong đó biến giải thích là biến giả

1.1.1. Bản chất của biến giả

- Ta thấy rằng các biến chất lượng thường có thuộc tính nào đó, chẳng hạn như nam hay nữ, doanh nghiệp nhà nước hay không phải là nhà nước... Trong phân tích hồi quy, người ta dùng một biến giả để lượng hoá được những thuộc tính như vậy.

- Biến giả có thể được sử dụng trong phạm vi hồi quy để giải thích cho sự kiện là: có những quan sát có thuộc tính gắn với một tập các tham số, còn các quan sát có thuộc tính khác sẽ gắn với một tập các tham số hồi quy khác. Ta sẽ thấy rằng, biến giả có thể sử dụng trong mô hình hồi quy như biến số lượng thông thường.

1.1.2. Mô hình hồi quy với biến giả

*) Mô hình một biến giả

- Giả sử một công ty sử dụng hai quá trình sản xuất (ký hiệu là A và B) để sản xuất ra một loại sản phẩm và giả sử sản phẩm đó là đại lượng ngẫu nhiên có phân phối chuẩn và có kỳ vọng khác nhau nhưng phương sai như nhau. Chúng ta có thể biểu thị quá trình sản xuất đó như một phương trình hồi quy như sau:

$$Y_i = \theta_0 + \theta_i D_i + \xi_i \qquad (4.1)$$

trong đó: Y_i là sản lượng gắn với quan sát thứ i, $i = 1, 2, ..., n$.

D_i là biến giả nhận một trong hai giá trị bằng 0 nếu sản lượng sản phẩm thu được từ quá trình sản xuất A và bằng 1 nếu sản lượng sản phẩm thu được từ quá trình sản xuất B.

Mô hình (4.1) chúng ta xét giống như mô hình hồi quy hai biến đã gặp trước đây, chỉ khác là biến số lượng X được thay thế bằng biến giả D. Căn cứ vào mô hình (4.1) chúng ta có thể biết được sản lượng trung bình do quá trình sản xuất A tạo ra có khác với sản lượng trung bình do quá trình sản xuất B tạo ra hay không?

Hệ số chặn θ_0 trong mô hình hồi quy tuyến tính đo sản lượng trung bình gắn với quá trình sản xuất B, trong khi đó θ_1 của phương trình hồi quy đo sự khác nhau về sản lượng sinh ra do việc thay đổi từ quá trình sản xuất B sang quá trình sản xuất A.

Kỳ vọng của Y tương ứng với biến D_i nhận giá 0 và 1 là:

$$E(Y_i \mid D_i = 0) = \theta_0$$

$$E(Y_i \mid D_i = 1) = \theta_0 + \theta_1$$

Ta sẽ dùng kiểm định giả thuyết H_0: $\theta_1 = 0$ để kiểm định về giả thuyết không có sự khác nhau về sản lượng do quá trình sản xuất A và B tạo ra. Ta có thể làm điều này giống như đã làm với mô hình hồi quy bội ở chương 3.

*) Mô hình với hai biến giả

- Ta thấy rằng thủ tục biến giả có thể mở rộng cho trường hợp có nhiều hơn hai thuộc tính. Chẳng hạn trong thí dụ trên ta giả thiết có 3 quá trình sản xuất để tạo ra sản phẩm và người ta hy vọng sẽ giải thích cho vấn đề là sản phẩm do các quá trình sản xuất khác nhau là không như nhau. Trong trường hợp này đưa vào hai biến giả là D_1 và D_2. Ta xét mô hình như sau:

$$Y_i = \theta_0 + \theta_1 D_{1i} + \theta_2 D_{2i} + \xi_i \qquad (4.2)$$

trong đó:

+) D_{1i} = 1 nếu sản lượng sản phẩm thu được từ quá trình sản xuất A.

D_{1i} = 0 nếu sản lượng sản phẩm thu được từ quá trình sản xuất khác.

+) D_{2i} = 1 nếu sản lượng sản phẩm thu được từ quá trình sản xuất B.

D_{2i} = 0 nếu sản lượng sản phẩm thu được từ quá trình sản xuất khác.

Như vậy 3 quá trình sản xuất này được biểu thị dưới dạng các kết hợp như dưới bảng 4.1 sau:

Bảng 4.1

Quá trình sản xuất	D_1	D_2
A	1	0
B	0	1
C	0	0

Bằng việc lấy kỳ vọng cho mỗi hàng chúng ta có thể giải thích kết quả hồi quy:

$$E(Y_i \mid D_{1i} = 1;\ D_{2i} = 0) = \theta_0 + \theta_1$$
$$E(Y_i \mid D_{1i} = 0;\ D_{2i} = 1) = \theta_0 + \theta_2$$
$$E(Y_i \mid D_{1i} = 0;\ D_{2i} = 0) = \theta_0$$

Hệ số chặn θ_0 của hồi quy biểu thị giá trị kỳ vọng của sản lượng do quá trình sản xuất C tạo ra, θ_1 đo sự thay đổi trung bình về sản lượng do việc chuyển từ quá trình sản xuất C sang quá trình sản xuất A và θ_2 đo sự thay đổi trung bình về sản lượng khi thay đổi từ quá trình sản xuất C sang quá trình sản xuất B.

Kiểm định giả thuyết H_0: $\theta_1 = 0$ có nghĩa là không có sự khác nhau giữa quá trình sản xuất A và quá trình sản xuất C. Giả thuyết H_0: $\theta_2 = 0$ cũng có ý nghĩa tương tự nhưng lại so sánh quá trình sản xuất C với B.

Ví dụ 4.1

Để xem xét kết quả sản lượng do hai quá trình sản xuất A và B có khác nhau hay không, người ta tiến hành lấy một mẫu gồm 10 quan sát được cho trong bảng 4.2. Ta sẽ phân tích kết quả hồi quy thu được.

Kết quả hồi quy như sau: $\hat{Y} = 18 + 3,2D$

$$\sqrt{\hat{D}(\hat{\theta}_0)} = 0,32 \qquad t_{\hat{\theta}_0} = 7,439$$
$$\sqrt{\hat{D}(\hat{\theta}_1)} = 0,44 \quad ; \quad t_{\hat{\theta}_1} = 57,74$$

Ta thấy rằng sản lượng trung bình một ca của quá trình sản xuất B được ước lượng bằng $\hat{\theta}_0 = 18$ nghìn kg, còn ước lượng cho sản lượng trung bình một của quá trình sản xuất A là $\hat{\theta}_0 + \hat{\theta}_1 = 18 + 3,2 = 21,2$ nghìn kg.

Bảng 4.2

Quá trình sản xuất (A: 1; B: 0)	Sản lượng trong một ca hoạt động (nghìn kg)
1	22,0
0	19,0
0	18,0
1	21,0
0	18,5
1	21,0
1	20,5
0	17,0
0	17,5
1	21,2

Qua tính toán ước lượng hệ số của mô hình trên ta thấy rằng sản lượng trung bình của hai quá trình đó là khác nhau.

Trong quá trình phân tích tính toán ta có thể rút ra được một số lưu ý:

1. Khi xem xét một vấn đề nào đó có hai thuộc tính chẳng hạn như nam hay nữ, quá trình sản xuất A hoặc B người ta dùng một biến giả, còn khi xem xét một vấn đề có ba thuộc tính người ta dùng hai biến giả, ..., tổng quát lên khi một vấn đề có N thuộc tính thì ta phải dùng $N-1$ biến giả. Số biến giả thấp hơn số thuộc tính để tránh hiện tượng đa cộng tuyến sẽ nói ở phần sau. Khi xem xét việc sản xuất ra sản phẩm nào đó, ta thấy sản phẩm đó có thể được tạo ra từ ba quá trình sản xuất A, B, C, ở đây ta đưa ra hai biến giả D_1, D_2. Việc đưa thêm biến giả D_3 ($D_3 = 1$ nếu sản lượng sản phẩm sản xuất từ quá trình C và $D_3 = 0$ nếu sản lượng sản phẩm sản xuất từ các quá trình khác) sẽ không cho thêm thông tin gì mà còn gây ra hiện tượng đa cộng tuyến hoàn hảo. Ở đây có hiện tượng đa cộng tuyến do:

$$D_3 = 1 - D_1 - D_2$$

2. Thuộc tính được gắn giá trị không được coi là thuộc tính cơ sở. Thuộc tính được gọi là cơ sở nếu như có sự so sánh thuộc tính này với thuộc tính khác trong cùng một hiện tượng nào đó. Như vậy trong mô hình sản xuất có ba thuộc tính đã nói ở trên thì quá trình C là thuộc tính cơ sở. Nếu ta đem ước lượng mô hình (4.2) với $D_1 = 0$; $D_2 = 0$ thì chỉ có quá trình sản xuất C và mô hình chỉ có $\hat{\theta}_0$.

3. Hệ số $\hat{\theta}_1$ gắn với biến giả D_1 được gọi là hệ số chặn chênh lệch, vì nó cho biết giá trị của số hạng chặn của thuộc tính nhận giá trị bằng 1 sẽ khác bao nhiêu với hệ số chặn của thuộc tính cơ sở.

1.2. Hồi quy với một biến lượng và một biến chất

Trong phần này ta sẽ xét mô hình hồi quy chỉ có một biến lượng và một biến chất, trong đó biến chất là biến giả có hai thuộc tính. Trường hợp có nhiều biến lượng và một biến chất thì quá trình tính toán sẽ tương tự.

Trong trường hợp này, mô hình hồi quy đơn giản như sau:

$$Y_i = \theta_0 + \theta_1 X_i + \theta_2 D_i + \xi_i \qquad i = 1, .., n \qquad (4.3)$$

với: Y_i là biến phụ thuộc; X_i là biến độc lập, D_i là biến giả là số quan sát trong mẫu.

Giả sử mô hình trên là mô hình hồi quy về sự phụ thuộc của tiền lương hàng tháng (Y_i) của một công nhân cơ khí vào bậc thợ của công nhân đó (X_i) và vào khu vực làm việc của công nhân đó (D_i) ở khu vực tư nhân (D_i = 1) hay quốc doanh (D_i = 0).

Trong mô hình đó thì biến lượng là bậc của công nhân đó và biến chất là khu vực làm việc của công nhân đó. Nếu ta giả thiết $E(\xi_i) = 0$ thì có thể biết được tiền lương của người công nhân làm ở khu vực quốc doanh khác tiền lương của người công nhân làm ở khu vực tư nhân hay không nếu các điều kiện khác không thay đổi.

Tiền lương bình quân của người công nhân làm ở khu vực quốc doanh là:

$$E(Y_i \mid X_i; D_i = 0) = \theta_0 + \theta_1 X_i \qquad (4.4)$$

Tiền lương bình quân của người công nhân làm ở khu vực tư nhân là:

$$E(Y_i \mid X_i; D_i = 1) = \theta_i + \theta_i X_i + \theta_2 \qquad (4.5)$$

Ta thấy rằng tiền lương trung bình của người công nhân làm ở khu vực quốc doanh khác với tiền lương trung bình của người công nhân làm ở khu vực tư nhân.

Chính nhờ ước lượng mô hình với biến giả giúp ta xác định được mức độ khác nhau của mô hình khi đặt biến chất ở những thuộc tính khác nhau.

§2. ĐA CỘNG TUYẾN

2.1. Đặt vấn đề

Trong mô hình phân tích hồi quy bội, ta đã giả thiết giữa các biến của mô hình không có quan hệ tương tác. Nhưng trong thực tế có thể xảy ra trường hợp ngược lại và gây ra hậu quả. Ta sẽ xét khi nào thì giả thiết trên bị vi phạm và tìm cách khắc phục.

Vậy một vấn đề cần quan tâm là khi nghiên cứu mô hình hồi quy ta sẽ xét không những về quan hệ của biến phụ thuộc Y_i và biến độc lập X_i mà còn nghiên cứu mối quan hệ giữa các biến độc lập với nhau.

Mô hình hồi quy tuyến tính cổ điển giả thiết các biến X_i không có tương quan với nhau, mỗi biến X_i chứa một thông tin về Y nhưng trong thực hành lại có sự phụ thuộc giữa các biến độc lập với nhau điều này dẫn đến hiện tượng đa cộng tuyến.

Giả sử ta phải ước lượng hàm hồi quy Y gồm k biến giải thích X_1, X_2, ..., X_k. Mô hình như sau:

$$Y = \theta_0 + \theta_1 X_1 + \theta_2 X_2 + ... + \theta_k X_k + \xi$$

Các biến X_1, X_2, ..., X_k được gọi là đa cộng tuyến nếu tồn tại các hằng số c_1, c_2, ..., c_k không đồng thời bằng 0 sao cho

$$c_0 + c_1 X_1 + c_2 X_2 + ... + c_k X_k = 0 \qquad (4.6)$$

trong đó c_1, c_2, ..., c_k là các hằng số không đồng thời bằng 0, c_0 là số ngẫu nhiên và do đó

$$c_0 + \sum_{j=1}^{k} c_i x_{ij} = 0 \,;\; i = 1, 2, ..., n$$

với n là số quan sát. Đây là đa cộng tuyến hoàn hảo.

Trong trường hợp này X có hạng $\leq k$ và không tồn tại ma trận nghịch đảo $(X^T X)^{-1}$. Trên thực tế $\left|X^T X\right| \approx 0$. Người ta có thể coi các X_1, X_2, ..., X_k có hiện tượng đa cộng tuyến. Khi đó ước lượng $\hat{\theta} = (X^T X)^{-1} X^T Y$ thường không ổn định và có phương sai rất lớn và khoảng tin cậy sẽ rất rộng.

Hiện tượng đa cộng tuyến thường biểu hiện qua các dấu hiệu sau:

+) Số các phần tử trên đường chéo chính V_{ij} của ma trận $V = (X^T X)^{-1}$ tỏ ra

rất lớn $j = 0, 1, 2, \ldots, k$.

+) Các hệ số tương quan tuyến tính mẫu của các cặp X_p, X_q là $r_{pq} = \dfrac{s_{pq}}{(s_{pp}s_{qq})^{1/2}}$ tỏ ra rất lớn. $p,q = 0, 1, 2, \ldots, k$

Ta hãy minh họa tư tưởng đa cộng tuyến bằng các đồ thị sau

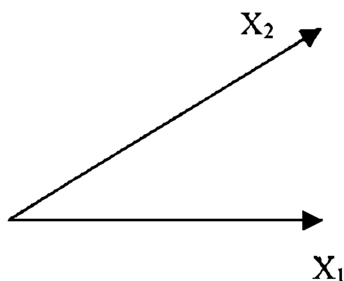
Các biến giải thích X — Biến phụ thuộc Y



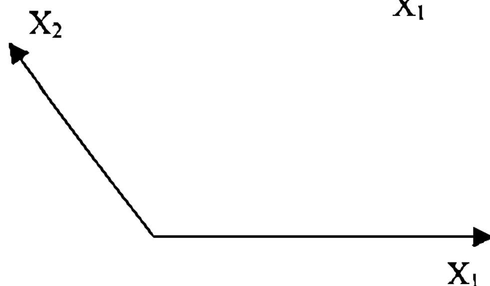
Các biến trực giao cho thông tin từ những nguồn độc lập: không có cộng tuyến



Thông tin đồng nhất, không có hồi quy: cộng tuyến hoàn toàn



Xuất hiện mức độ cộng tuyến nào đó giữa các biến X, nhưng vấn đề hồi quy phụ thuộc vào bậc của cộng tuyến



Tương quan ngược chiều giữa các biến: cộng tuyến mạnh

2.2. Ước lượng khi có đa cộng tuyến hoàn hảo

Dưới đây, chúng ta sẽ chỉ ra rằng khi có đa cộng tuyến hoàn hảo thì các hệ số hồi quy là không còn xác định, còn các sai số tiêu chuẩn là vô hạn. Để đơn giản về mặt trình bày, chúng ta xét mô hình hồi quy ba biến:

$$Y = \theta_0 + \theta_1 X_1 + \theta_2 X_2 + \xi$$

Đặt:

$$y_i = Y_i - \overline{Y}, \quad y_i = X_i - \overline{X}$$

$$\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i, \quad \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

thì mô hình hồi quy ba biến có thể viết lại như sau:

$$y_i = \theta_1 x_{1i} + \theta_2 x_{2i} + \gamma_i$$

Theo phương pháp BPCT ta thu được ước lượng cho các hệ số θ_1, θ_2 như sau:

$$\hat{\theta}_1 = \frac{(\sum_{i=1}^{n} y_i x_{1i})(\sum_{i=1}^{n} x_{2i}^2) - (\sum_{i=1}^{n} y_i x_{2i})(\sum_{i=1}^{n} x_{1i} x_{2i})}{(\sum_{i=1}^{n} x_{1i}^2)(\sum_{i=1}^{n} x_{2i}^2) - (\sum_{i=1}^{n} x_{1i} x_{2i})^2}$$

$$\hat{\theta}_2 = \frac{(\sum_{i=1}^{n} y_i x_{2i})(\sum_{i=1}^{n} x_{1i}^2) - (\sum_{i=1}^{n} y_i x_{1i})(\sum_{i=1}^{n} x_{1i} x_{2i})}{(\sum_{i=1}^{n} x_{1i}^2)(\sum_{i=1}^{n} x_{2i}^2) - (\sum_{i=1}^{n} x_{1i} x_{2i})^2}$$

Giả sử $X_{2i} = cX_{1i}$ trong đó c là hằng số khác 0, thay điều kiện này ước lượng cho $\hat{\theta}_1$ ta được:

$$\hat{\theta}_2 = \frac{c(\sum_{i=1}^{n} y_i x_{1i})(\sum_{i=1}^{n} x_{1i}^2) - (\sum_{i=1}^{n} y_i x_{1i})c(\sum_{i=1}^{n} x_{1i}^2)}{(\sum_{i=1}^{n} x_{1i}^2)c^2(\sum_{i=1}^{n} x_{1i}^2) - c^4(\sum_{i=1}^{n} x_{2i}^2)^2} = 0$$

$$\hat{\theta}_1 = \frac{(\sum_{i=1}^{n} y_i x_{1i})(c^2 \sum_{i=1}^{n} x_{1i}^2) - c(\sum_{i=1}^{n} y_i x_{1i})c\sum_{i=1}^{n} x_{1i}^2}{(\sum_{i=1}^{n} x_{1i}^2)(c^2 \sum_{i=1}^{n} x_{1i}^2) - c^4(\sum_{i=1}^{n} x_{1i}^2)^2} = 0$$

Tương tự như thế đối với $\hat{\theta}_2$ ta cũng được:

$$\hat{\theta}_2 = \frac{c(\sum_{i=1}^{n} y_i x_{1i})(\sum_{i=1}^{n} x_{1i}^2) - (\sum_{i=1}^{n} y_i x_{1i})c(\sum_{i=1}^{n} x_{1i}^2)}{(\sum_{i=1}^{n} x_{1i}^2)c^2(\sum_{i=1}^{n} x_{1i}^2) - c^4(\sum_{i=1}^{n} x_{1i}^2)^2} = 0$$

Ví dụ 4.2:

Nếu $X_{2i} = cX_{1i}$ thay điều kiện này vào mô hình ta có:

$$y_i = \theta_1 x_{1i} + \theta_2 x_{2i} + \gamma_i = (\theta_1 + c\theta_2)x_{1i} + \gamma_i$$

Đặt $\theta_1 + c\theta_2 = \alpha$ mô hình khi đó trở thành $y_i = \alpha x_{1i} + \gamma_i$. Áp dụng phương pháp BPCT ta được:

$$\hat{\alpha} = \frac{\sum_{i=1}^{n} y_i x_{1i}}{\sum_{i=1}^{n} x_{1i}^2}$$

Như vậy dù ước lượng được cho α là $\hat{\alpha}$ nhưng ta cũng không thể ước lượng được cho các hệ số hồi quy của phương trình hồi quy ba biến ban đầu do ta chỉ có một phương trình mà có tới hai ẩn số.

Trong trường hợp đa cộng tuyến hoàn hảo, chúng ta không thể nhận được lời giải cho các hệ số hồi quy mà chỉ nhận được lời giải duy nhất cho tổ hợp tuyến tính của các hệ số này. Chú ý rằng trong trường hợp đa cộng tuyến hoàn hảo thì phương sai và các sai số tiêu chuẩn của các ước lượng $\hat{\theta}_1, \hat{\theta}_2$ bằng 0.

2.3. Ước lượng trong trường hợp có đa cộng tuyến không hoàn hảo

Đa cộng tuyến hoàn hảo chỉ là trường hợp đặc biệt xảy ra. Trong các số liệu quan sát có liên quan đến chuỗi thời gian sẽ được nói ở mục sau, thường xảy ra đa cộng tuyến không hoàn hảo.

Xét mô hình hồi quy ba biến như sau:

$$Y = \theta_0 + \theta_1 X_1 + \theta_2 X_2 + \xi$$

Chúng ta sẽ giả thiết rằng giữa X_1 và X_2 có cộng tuyến hoàn hảo theo nghĩa: $x_{2i} = cx_{1i} + \varepsilon_i$ trong đó c là hằng số khác không, ε_i là nhiễu ngẫu nhiên hay sai số sao cho $\sum_{i=1}^{n} x_{1i}\varepsilon_i = 0$, trong trường hợp này ta vẫn có thể ước lượng được các hệ số của mô hình.

2.4. Hậu quả của đa cộng tuyến

Trong trường hợp tồn tại đa cộng tuyến hoàn hảo ta gặp một số tình huống sau:

*) Phương sai và hiệp phương sai của các ước lượng BPCT lớn

Chúng ta có:

$$D(\hat{\theta}_1) = \frac{\sigma^2}{\sum_{i=1}^{n}(1 - r_{12}^2)x_{1i}^2}; \quad D(\hat{\theta}_2) = \frac{\sigma^2}{\sum_{i=1}^{n}(1 - r_{12}^2)x_{2i}^2}$$

và $$\operatorname{cov}(\hat{\theta}_1, \hat{\theta}_2) = \frac{-r_{12}\sigma^2}{(1 - r_{23}^2)\sqrt{\sum_{i=1}^{n} x_{1i}^2 \sum_{i=1}^{n} x_{2i}^2}}$$

trong đó r_{12} là hệ số tương quan giữa X_1 và X_2.

Từ các công thức trên chúng ta nhận thấy khi r_{12} tăng dần đến 1 (nghĩa là cộng tuyến tăng) thì phương sai của hai ước lượng này tăng đến vô hạn.

*) Khoảng tin cậy rộng lớn

*) Tỷ số t mất ý nghĩa

Khi ta kiểm định giả thuyết $H_0 : \theta_1 = 0$, chúng ta sử dụng tỷ số t

$$t = \frac{\hat{\theta}_1}{\sqrt{D(\hat{\theta}_1)}}$$

và đem so sánh giá trị t đã được ước lượng với giá trị tới hạn t. Nhưng khi có đa cộng tuyến gần hoàn hảo thì sai số tiêu chuẩn ước lượng sẽ có giá trị rất cao. Vì vậy làm cho tỷ số t nhỏ đi. Kết quả sẽ làm tăng khả năng chấp nhận giả thuyết $H_0 : \theta_1 = 0$

*) R^2 cao nhưng chỉ số t ít có ý nghĩa.

Để giả thích điều này chúng ta xét mô hình k biến

$$Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + u_i$$

Trong trường hợp này có đa cộng tuyến gần hoàn hảo, như đã chỉ ra ở trên ta có thể tìm được một hoặc một số hệ số góc riêng là không có ý nghĩa về mặt bằng kiểm định F, chúng ta có thể bác bỏ giả thuyết H_0: $\theta_1 = \theta_2 = \ldots = \theta_k = 0$.

Mâu thuẫn này cũng có tín hiệu của đa cộng tuyến.

*) Các ước lượng bình phương bé nhất và sai số tiêu chuẩn của chúng trở nên rất nhạy cảm đối với những thay đổi nhỏ của bộ số liệu.

*) Dấu của các ước lượng của các hệ số của q có thể sai.

Khi có đa cộng tuyến gần hoàn hảo thì có thể thu được các ước lượng của các hệ số quy hồi trái với điều chúng ta mong đợi. Chẳng hạn lý thuyết kinh tế cho rằng đối với hàng hoá bình thường khi thu nhập tăng, cầu hàng hoá tăng, nghĩa là khi hồi quy thu nhập là một trong các biến, còn các hàng hoá là biến được giải thích, nếu xảy ra hiện tượng đa cộng tuyến gần hoàn hảo thì ước lượng của hệ số của biến thu nhập có thể mang dấu âm điều này mâu thuẫn với điều ta mong đợi.

*) Thêm vào hay bớt đi các biến cộng tuyến với các biến khác, mô hình sẽ thay đổi về độ lớn của các ước lượng hoặc dấu của chúng.

Tóm lại, triệu chứng của đa cộng tuyến ta đã nói ở trên là tăng sai số tiêu chuẩn. Sai số tiêu chuẩn cao hơn có ngụ ý rằng sự biến thiên của hệ số hồi quy từ mẫu này sang mẫu khác cao hơn, do đó một sự thay đổi nhỏ trong số liệu hoặc trong mô hình hồi quy (như thêm vào hoặc bớt đi một biến) sẽ gây ra sự thay đổi lớn của các hệ số.

Như vậy chúng ta đã biết được một số hậu quả của đa cộng tuyến. Nhưng dù hậu quả thế nào đi chăng nữa thì điều quan trọng là làm thế nào để thấy được sự tồn tại của nó để ta có thể ngăn ngừa những hậu quả tai hại trong thủ tục ước lượng sử dụng mô hình hồi quy để dự đoán, điều khiển hoặc hiểu quá trình liên quan tới nó. Dưới đây là một số phương pháp để phát hiện đa cộng tuyến và mô tả những triệu chứng chủ yếu của nó.

2.5. Phát hiện sự tồn tại của đa cộng tuyến

*) R^2 cao nhưng chỉ số thấp.

Khi R^2 cao (thường $R^2 > 0,8$) mà tỷ số t lại có giá trị thấp đây chính là dấu hiệu của đa cộng tuyến.

*) Tương quan giữa các biến giải thích cao.

Nếu hệ số tương quan giữa các biến giải thích cao (vượt 0,8) thì có khả năng tồn tại đa cộng tuyến. Tuy nhiên tiêu chuẩn này thường không chính xác.

*) Xem xét tương quan riêng.

Trong một số trường hợp chúng ta sử dụng các hệ số tương quan riêng cũng có thể đoán được sự tồn tại của đa cộng tuyến, nhưng nó cũng không đảm bảo rằng sẽ cung cấp cho ta hướng chính xác trong việc phát hiện đa cộng tuyến.

*) Hồi quy phụ

Một cách có thể tin cậy được để đánh giá mức độ của đa cộng tuyến là hồi quy phụ. Hồi quy phụ là hồi quy trong đó mỗi một biến giải thích X_j ; j = 1, 2, ..., k sẽ theo các biến giải thích còn lại. R^2 được tính từ hồi quy này ta có ký hiệu là R_j^2 và có mối quan hệ với F_j. Trong đó F_j tuân theo quy luật phân phối Fisher với $k - 2$ và $n - k + 1$ bậc tự do (n là cỡ mẫu, k là số biến giải thích kể cả hệ số chặn trong mô hình). Nếu F_j tính được vượt điểm tới hạn $F_j(k - 2, n - k + 1)$ ở mức ý nghĩa đã cho thì có nghĩa là X_j có liên hệ tuyến tính với các biến X còn lại trong mô hình. Trong trường hợp đó ta giữ lại các biến trong mô hình. Nếu F_j có nghĩa về mặt thông kê chúng ta vẫn phải quyết định liệu biến X_j nào sẽ bị loại khỏi mô hình. Một trở ngại của kỹ thuật hồi quy phụ là gánh nặng tính toán, nhưng ngày nay nhiều chương trình máy tính đã có thể đảm đương được công việc tính toán này. Ngoài ra chúng ta cũng có thể sử dụng một số phương pháp sau để phát hiện ra sự tồn tại của đa cộng tuyến đó là: Nhân tử phóng đại phương sai, Độ đo Theil.

2.6. Biện pháp khắc phục hiện tượng đa cộng tuyến

Để khắc phục hiện tượng đa cộng tuyến, chúng ta có thể sử dụng một số phương pháp sau:

*) Sử dụng thông tin tiên nghiệm

Một trong các cách tiếp cận để giải quyết vấn đề đa cộng tuyến là phải tận dụng thông tin tiên nghiệm hoặc thông tin từ các nguồn khác để ước lượng các hệ số riêng. Thông tin tiên nghiệm có thể giúp ta giảm số biến độc lập trong mô hình.

*) Thu thập số liệu hoặc lấy thêm mẫu mới

Vì đa cộng tuyến là đặc trưng của mẫu nên có thể có mẫu khác liên quan đến cùng các biến trong mẫu ban đầu mà cộng tuyến có thể không nghiêm trọng nữa. Điều này chỉ có thể làm được khi chi phí cho việc lấy mẫu khác có thể chấp nhận được trong thực tế. Đôi khi chỉ cần thu thập thêm số liệu, tăng cỡ mẫu có thể giảm tính nghiêm trọng của đa cộng tuyến.

*) Bỏ biến

Khi có hiện tượng đa cộng tuyến thì cách đơn giản nhất là bỏ biến cộng tuyến ra khỏi mô hình.

Ngoài các biện pháp trên, ta có thể sử dụng một số phương pháp như:

*) Sử dụng sai phân cấp một

*) Giảm tương quan trong hồi quy đa thức.

Để khắc phục hiện tượng đa cộng tuyến có thể làm theo cách sau:

Đặt r_{0q} là hệ số tương quan tuyến tính mẫu giữa Y và X_q, $q = 0, 1, ..., k$. Cụ thể hơn :

$$r_{0q} = \frac{s_{0q}}{(s_{qq}s_{00})^{1/2}}$$

trong đó : $s_{00} = s_y^2$; $s_{0q} = \frac{1}{n}\sum_{i=1}^{n} y_i x_{ip} - \bar{y} \times \bar{x}_p$

Khi đó nếu thấy $|r_{pq}| \geq 0,7$ thì ta sẽ loại biến X_q ra khỏi mô hình nếu $|r_{0p}| > |r_{0q}|$ và sẽ loại biến X_p ra khỏi mô hình nếu $|r_{0p}| < |r_{0q}|$.

§3. TỰ TƯƠNG QUAN

3.1. Khái niệm chuỗi thời gian

Hầu hết những thủ tục thống kê cổ điển dùng những số liệu xuất phát từ một loạt các quan sát độc lập, thường không quan tâm đến thứ tự quan sát diễn ra.trong mẫu quan sát. Tuy nhiên, với mẫu quan sát dưới dạng chuỗi thời gian thì tình hình không như thế nữa.

Một chuỗi thời gian là một dãy các giá trị quan sát $x_1, x_2, ..., x_n$ được xếp thứ tự theo thời gian, x_1 là giá trị quan sát tại thời điểm đầu tiên, x_2 là giá trị quan sát tại thời điểm thứ hai, ..., x_t là giá trị quan sát tại thời điểm t, ..., x_n là giá trị quan sát tại thời điểm thứ n và cũng là thời điểm cuối cùng. Chẳng hạn, mức nước sông Hồng đo từng giờ ở trạm đo Trung Hà, số lượng hành khách đi hàng ngày trên chuyến tàu Bắc Nam, bản theo dõi điện tâm đồ của một bệnh nhân tim mạch...

Trong kinh tế, các số liệu có được hầu hết ở dạng chuỗi thời gian. Khi nghiên cứu về các chuỗi thời gian, ta thấy thường xuất hiện sự phụ thuộc, ảnh hưởng lẫn nhau về số liệu quan sát tức là tạo nên hiện tượng tự tương quan.

Vậy nên trong chương này ta sẽ tìm hiểu rõ hơn về hiện tượng tự tương quan trong mô hình chuỗi thời gian.

3.2. Tự tương quan là gì?

Thuật ngữ tự tương quan có thể là sự tương quan giữa các thành phần của chuỗi các quan sát được sắp xếp theo đơn vị thời gian hoặc không gian.

Trong phạm vi hồi quy, mô hình hồi quy tuyến tính cổ điển giả thiết không có sự tự tương quan giữa các biến nhiễu $cov(\xi_i, \xi_j) = 0$ $(i \neq j)$.

Hay nói một cách khác, trong mô hình hồi quy tuyến tính cổ điển thì thành phần nhiễu gắn với mọi quan sát này không bị ảnh hưởng bởi thành phần nhiễu ứng với các quan sát khác.

Tuy nhiên trong thực tế có xảy ra hiện tượng mà các thành phần của quan sát lại có thể phụ thuộc lẫn nhau, nghĩa là $cov(\xi_i, \xi_j) \neq 0$ $(i \neq j)$.

3.3. Nguyên nhân của sự tự tương quan

3.3.1. Nguyên nhân khách quan

*) Quán tính

Nét nổi bật của hầu hết các chuỗi thời gian trong kinh tế là quán tính. Trong hồi quy của chuỗi thời gian các quan sát kế tiếp có nhiều khả năng phụ thuộc lẫn nhau.

*) Hiện tượng mạng nhện

Hiện tượng mạng nhện là hiện tượng biến phụ thuộc ở thời điểm t bị ảnh hưởng bởi biến giải thích ở thời điểm $t - 1$ trước đó.

Khi xuất hiện hiện tượng này thì hầu hết các giá trị sau bị ảnh hưởng vào giá trị trước nên các nhiễu có ảnh hưởng lẫn nhau và gây nên hiện tượng tự tương quan.

*) Hiện tượng trễ

Hiện tượng biến phụ thuộc ở thời kỳ t phụ thuộc vào chính biến đó ở thời kỳ $t - 1$ và các biến độc lập khác.

3.3.2. Nguyên nhân chủ quan

*) Xử lý số liệu

Trong phân tích thực nghiệm, số liệu thô thường được xử lý bằng cách lấy trung bình. Việc lấy trung bình này làm trơn các số liệu và làm giảm sự dao động dẫn tới sai số hệ thống trong các biến nhiễu ngẫu nhiên và gây ra tự tương quan. Hay khi xử lý số liệu bằng phép nội suy và ngoại suy cũng có thể gây ra sai số hệ thống.

*) Sai số do lập mô hình

Đây là nguyên nhân thuộc về lập mô hình, có hai sai lầm gây ra tự tương quan:

- Không đưa đủ các biến vào trong mô hình.
- Dạng hàm mô hình hoá sai.

3.4. Ước lượng BPCT khi có sự tự tương quan

Giả sử tất cả các giả thiết của mô hình hồi quy tuyến tính cổ điển là thỏa mãn, trừ giả thiết là có tự tương quan giữa các nhiễu ξ_i, ξ_j nghĩa là:

$$\text{cov}(\xi_i, \xi_j) \neq 0 \ (i \neq j)$$

Xét mô hình hồi quy tuyến tính đơn giản:

$$Y_t = \theta_1 + \theta_2 X_t + \xi_t \qquad (4.7)$$

Do giả thiết $\text{cov}(\xi_i, \xi_j) \neq 0$ $(i \neq j)$ ta có thể giả thiết nhiễu sản sinh theo cách sau:

$$\xi_t = \rho\xi_{t-1} + \varepsilon_t (-1 < \rho < 1) \qquad (4.8)$$

$$\begin{cases} E(\varepsilon_t) = 0 \\ \text{cov}(\varepsilon_t . \varepsilon_{t-s}) = 0 (s \neq 0) \\ \text{var}(\varepsilon_t) = \sigma^2 \end{cases} \qquad (4.9)$$

Khi đó ta nói rằng ξ_t có tự tương quan dương, âm, không có tự tương quan nếu $\rho > 0$, $\rho < 0$, $\rho = 0$.Theo phương pháp BPCT ta tính được các giá trị thực nghiệm: $\hat{\theta}_2$

$$\hat{\theta}_2 = \frac{\sum_{t=1}^{n} x_t y_t}{\sum_{t=1}^{n} x_t^2} \quad (4.10)$$

$$D(\hat{\theta}_2) = \frac{\sigma^2}{\sum_{t=1}^{n} x_t^2} + 2\frac{\sigma^2}{\sum_{t=1}^{n} x_t^2}\left[\rho\frac{\sum_{t=1}^{n-1} x_t x_{t+1}}{\sum_{t=1}^{n} x_t^2} + \rho^2\frac{\sum_{t=1}^{n-2} x_t x_{t+2}}{\sum_{t=1}^{n} x_t^2} + \ldots + \rho^{n-1}\frac{x_1 x_n}{\sum_{t=1}^{n} x_t^2}\right]$$

(4.11)

$$x_t = X_t - \bar{X}$$
$$y_t = Y_t - \bar{Y}$$

3.5. Hậu quả của việc sử dụng phương pháp BPCT khi có tự tương quan

- Ước lượng BPCT thông thường vẫn là ước lượng tuyến tính không chệch, nhưng chúng không phải là ước lượng hiệu quả.

- Phương sai ước lượng được của các ước lượng BPCT thông thường là ước lượng chệch.Việc tính phương sai và sai số tiêu chuẩn của các ước lượng BPCT thường đưa đến ước lượng quá thấp phương sai thực và sai số tiêu chuẩn. Hậu quả sẽ ngộ nhận một hệ số nào đó khác không có ý nghĩa về mặt thống kê nhưng thực tế lại không phải như vậy.

- Kiểm định t và F không đáng tin cậy...

3.6. Phát hiện có sự tự tương quan

3.6.1. Kiểm định các đoạn mạch

*) Giới thiệu qua về kiểm định đoạn mạch.

- Một đoạn mạch là một dãy các phần tử giống nhau mà ở sát trước hoặc sát sau là các phần tử khác chúng hoặc không có phần tử nào. Chiều dài của đoạn mạch là số phần tử của đoạn mạch đó.

- Kiểm định các đoạn mạch là một phép kiểm định thống kê cho phép giả thiết một dãy các ký hiệu, các khoản mục, các số liệu có phải là kết quả của một quá trình mang tính ngẫu nhiên hay không?

Ví dụ 4.3:

Trong mô hình chuỗi thời gian về số lượng các viên thuốc bổ tổng hợp đã xây dựng, các sai số thu được lần lượt là:

-23	30	12	-10	-5	-17	-22	57	43	-23	
31	42	50	61	-28	-52	10	34	28	55	60
32	88	-75	-22	-56	-89	34	-20	-2	-5	29
12	45	77	78	91	25	60	-25	45	42	30
-59	-60	-40	-75	-25	-34	-66	-90	10	-20	

Viết lại dãy số liệu dưới dạng đoạn mạch:

− + + − − − − + + − + + + + − − + + + + + + + − − −

− − − − − + + + + + + + + − + + + − − − − − − − − + −

Để xác định có bao nhiêu đoạn mạch có thể chấp nhận là quá trình ngẫu nhiên, ta dùng một quy luật phân phối xác suất.

*) Nội dung của kiểm định đoạn mạch

- Với n là tổng số quan sát; n_1, n_2 lần lượt là các sai số dương và âm. N là số đoạn mạch.

- Giả thuyết kiểm định

Giả thuyết H_0: sai số có tương quan.

Đối thuyết H_1: các sai số không có tương quan.

Với giả thiết $n_1 \geq 10$; $n_2 \geq 10$

- Kỳ vọng cho số đoạn mạch là: $E(N) = \dfrac{2n_1.n_2}{n_1 + n_2} + 1$ (4.12)

- Phương sai cho N là: $\sigma_N^2 = \dfrac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}$ (4.13)

- Độ lệch chuẩn cho N: $\sigma_N = \sqrt{\dfrac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}}$ (4.14)

Nếu giả thuyết về tính ngẫu nhiên có thể chấp nhận được chúng ta tìm khoảng ước lượng cho số đoạn mạch N với độ tin cậy $1 - \alpha$ cho trước

$$\left[E(N) - \sigma_N.z_b;\ E(N) + \sigma_N.z_b\right];\quad z_b = z_{1-\frac{\alpha}{2}}$$

Ví dụ 4.4: Trong ví dụ 4.2 ta thấy: có tất cả N = 15 đoạn mạch; n_1 = 27; n_2 = 26.

Áp dụng các công thức (4.12); (4.14) để tính

$$E(N) = \frac{2n_1.n_2}{n_1 + n_2} + 1 = \frac{2.26.26}{27 + 26} + 1 \approx 27,4$$

$$\sigma_N = \sqrt{\frac{2n_1 n_2(2n_1 n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}} = \sqrt{\frac{2.26.27(2.26.27 - 26 - 27)}{(26 + 27)^2(26 + 27 - 1)}} = 3,6$$

Khoảng tin cậy cho với độ tin cậy 95% cho N là $\left[E(N) - \sigma_N.z_b; E(N) + \sigma_N.z_b\right]$. Tra bảng Laplace với độ tin cậy 95% thì z_b = 1,96. Khoảng tin cậy

[27,49 – 1,96.3,6; 27,49 + 1,96] = [20,43; 34,55]. Mà N = 10 vậy $N \notin$ [20,43; 34,55] $\Rightarrow$ Không có tự tương quan.

3.6.2. Kiểm định χ^2 về tính độc lập của các sai số

*) Bảng tiếp liên

Bảng tiếp liên là bảng trong đó:

- Các cột ứng với sai số âm và dương tại t

- Các dòng ứng với sai số dương và âm tại t – 1

Ở trong mục kiểm định này, ta dùng bảng tiếp liên có hai dòng và hai cột để ghi các giá trị tại thời điểm t và t – 1. Ta sẽ thêm một dòng và một cột để tính tổng.

- Trong mỗi ô tính A_{ij}, E_{ij} tương ứng là tần số thực tế và tần số lý thuyết của các ô (i,j)

- Cột cuối cùng là tổng theo dòng $R_i = \sum_{j=1}^{2} A_{ij};\ i = 1,2$

- Dòng cuối cùng là tổng theo cột $c_j = \sum_{j=1}^{2} A_{ij};\ i = 1,2$

- Ô cuối cùng ghi kích thước mẫu.

Bảng 4.3: Bảng tiếp liên 3 dòng, 3 cột

	Sai số dương tại t	Sai số âm tại t	R_i
Sai số dương tại $t-1$	A_{11} E_{11}	A_{12} E_{12}	R_1
Sai số âm tại $t-1$	A_{21} E_{21}	A_{22} E_{22}	R_2
c_j	c_1	C_2	n

*) Kiểm định

- Giả thuyết H_0 là các dòng và các cột có tương quan với nhau
- Đối thuyết H_1 là các hàng và các cột không tương quan với nhau

Tiêu chuẩn kiểm định χ^2 cho tập hợp các giả thiết này

Thống kê: $$\chi^2 = \sum_{i=1}^{2}\sum_{j=1}^{2}\frac{\left(A_{ij} - E_{ij}\right)^2}{E_{ij}} \quad (4.15)$$

Nếu giả thuyết H_0 đúng tức là các sai số là độc lập. Khi đó thống kê χ^2 có phân phối χ^2 với 1 bậc tự do.

H_0 đúng thì các dòng và cột độc lập với nhau và khi đó $E_{ij} = nP_{ij} = nP_iP_j$, trong đó P_{ij} là xác suất để xảy ra đồng thời sự kiện i và j, còn P_i, P_j là các xác suất để xảy ra sự kiện i và j

Từ bảng ta thấy $P_i = \frac{R_i}{n}$; $P_j = \frac{c_j}{n}$

$$E_{ij} = nP_iP_j = n\frac{R_i}{n}\frac{c_j}{n} = \frac{R_ic_j}{n}$$

3.6.3. Kiểm định d.Durbin-Watson

*) Định nghĩa thống kê Durbin-Watson

- Thống kê Durbin-Watson được định nghĩa như sau:

$$DW = \frac{\sum_{t=2}^{n} (\hat{\xi}_t - \hat{\xi}_{t-1})^2}{\sum_{t=1}^{n} \hat{\xi}_t^2} \qquad (4.16)$$

Vậy nên thống kê Durbin-Watson là tỉ số của tổng bình phương hiệu các sai số kế tiếp nhau với tổng bình phương các sai số.

- Các giá trị làm cơ sở cho thống kê Durbin-Watson

+) Mô hình hồi quy phải bao gồm số hạng chặn trục Y là ξ_0.

+) Các biến giải thích là phi ngẫu nhiên hoặc cố định trong phép lấy mẫu độc lập.

+) Các nhiễu ξ_t được sản sinh ra từ số ξ_{t-1}. Tức là quá trình tự hồi quy: $\xi_t = \rho\xi_{t-1} + \alpha_t$

+) Mô hình đem ra xét không phải là quá trình tự hồi quy.

+) Không có quan sát bị mất dữ liệu.

6.3.4. Kiểm định Durbin-Watson

$$DW = \frac{\sum_{t=2}^{n} (\hat{\xi}_t - \hat{\xi}_{t-1})^2}{\sum_{t=1}^{n} \hat{\xi}_t^2} = \frac{\sum_{t=2}^{n} \hat{\xi}_t^2 - 2\sum_{t=2}^{n} \hat{\xi}_t \hat{\xi}_{t-1} + \sum_{t=2}^{n} \hat{\xi}_{t-1}^2}{\sum_{t=1}^{n} \hat{\xi}_t^2}$$

Nếu giả thiết $\sum_{t=2}^{n} \hat{\xi}_t^2 = \sum_{t=2}^{n} \hat{\xi}_{t-1}^2$ thì $DW = 2\left(1 - \frac{\sum_{t=2}^{n} \hat{\xi}_t \hat{\xi}_{t-1}}{\sum_{t=1}^{n} \hat{\xi}_t^2}\right)$.

Định nghĩa $\hat{\rho} = \frac{\sum_{t=2}^{n} \hat{\xi}_t \hat{\xi}_{t-1}}{\sum_{t=1}^{n} \hat{\xi}_t^2}$

Ta sẽ có $DW = 2(1 - \hat{\rho})$ (4.17)

tuân theo phân phối Durbin-Watson với mức ý nghĩa α, k biến độc lập, kích thước mẫu n ta sẽ tìm được $d_L(k,n,\alpha) < d_U(k,n,\alpha)$. Khi so sánh Durbin-

Watson với d_L, d_U ta đưa ra các kết luận sau:

+) Nếu $0 \le DW \le d_L$ các ξ_t có tự tương quan cấp 1 dương.

+) Nếu $d_L \le DW \le d_U$ thì chưa kết luận được gì.

+) Nếu $d_U < DW < 4 - d_U$ thì các ξ_t có tự tương quan dương cấp 1.

+) Nếu $4 - d_U \le DW \le 4 - d_L$ thì không thể kết luận được.

+) Nếu $4 - d_L < DW \le 4$ thì các ξ_t có tự tương quan cấp 1 âm.

Ví dụ 4.5: Giả sử chúng ta tiến hành hồi quy theo phương pháp bình phương nhỏ nhất với $n = 20$ số liệu, số biến độc lập $k = 3$. Giá trị tính toán của thống kê Durbin-Watson là $DW = 3,1$; $\alpha = 0,01$.

Tra bảng Durbin-Watson:

$$d_L(3; 20; 0,01) \approx 0,77; \quad d_U(3; 20; 0,01) = 1,41$$

$$\Rightarrow 4 - d_L = 4 - 0,77 = 3,23; \quad 4 - d_U = 4 - 1,41 = 2,59$$

$$\Rightarrow 4 - d_U = 2,59 < DW < 4 - d_L = 3,23$$

Vậy nên không thể có kết luận gì được.

3.7. Chọn mô hình và kiểm định việc chỉ định mô hình

3.7.1. Đặt vấn đề

Khi đưa ra mô hình, ta ngầm định mô hình đã chọn là đúng. Nhưng có thể xảy ra sai lầm trong việc chỉ định mô hình, có nghĩa là thay cho việc ước lượng một mô hình đúng, ta lại ước lượng một mô hình không đúng. Nếu việc này xảy ra thì hậu quả sẽ như thế nào?Chúng ta có thể gặp những sai lầm chỉ định nào? Làm thế nào để phát hiện các sai lầm chỉ định đó.

Trước khi trả lời các câu hỏi này ta phải hiểu thế nào là một mô hình tốt và làm thế nào để có thể đưa ra một mô hình tốt. Do đó phải đưa ra các tiêu chuẩn để giải đáp cho câu hỏi: "Thế nào là một mô hình đúng?".

Trước tiên, chúng ta hãy xem xét thế nào là một mô hình tốt. Một mô hình tốt là mô hình có các tiêu chuẩn sau:

- Tính kiệm: chọn mô hình càng đơn giản càng tốt.

- Tính đồng nhất: với một tập hợp dữ liệu đã cho, các tham số ước lượng phải có giá trị đồng nhất.

- Tính thích hợp: một mô hình hồi quy tốt là mô hình có R^2 cao.

- Tính vững về mặt lý thuyết: trong việc xây dựng mô hình, ta phải có một cơ sở lý thuyết nào đó, bởi phép đo không có lý thuyết thường dẫn đến kết quả tồi.

- Khả năng dự đoán: khả năng sự đoán của mô hình phải phù hợp với thực tế.

3.7.2.Các sai lầm chỉ định

*) Bỏ sót một số biến thích hợp:

Giả sử mô hình đúng biểu thị mối liên hệ kinh tế giữa biến phụ thuộc Y và các biến giải thích X_1, X_2 có dạng:

$$Y_t = \theta_0 + \theta_1 X_{1t} + \theta_2 X_{2t} + \xi_t \qquad (4.18)$$

trong đó: θ_0, θ_1, θ_2 là các hệ số; ξ_t là sai số ngẫu nhiên.

Nhưng vì lý do nào đó ta ước lượng mô hình

$$Y_t = \alpha_0 + \alpha_1 X_{1t} + \lambda_t \qquad (4.19)$$

Mô hình ước lượng được thiếu biến X_2, sẽ gây tác hại:

1. Nếu X_2 tương quan với biến X_1 thì $\hat{\alpha}_0, \hat{\alpha}_1$ lần lượt là ước lượng chệch của θ_0, θ_1. Tức là $E\hat{\alpha}_0 \neq \theta_0$, $E\hat{\alpha}_1 \neq \theta_1$.

2. $\hat{\alpha}_0, \hat{\alpha}_1$ không phải là ước lượng vững.

3. Nếu X_1 và X_2 không tương quan khi đó $E\hat{\alpha}_1 = \theta_1$ khi đó $\hat{\alpha}_1$ chính là ước lượng không chệch của θ_1 và đồng thời nó cũng là ước lượng vững. Trong khi đó $\hat{\alpha}_0$ là ước lượng chêch của θ_0.

4. Ước lượng cho phương sai của sai số trong (4.18) sẽ khác ước lượng cho phương sai của sai số trong (4.19)

5. Ước lượng cho phương sai $D\hat{\alpha}_1 = \dfrac{\sigma^2}{\sum x_{1t}^2}; x_{1t} = X_{1t} - \overline{X}$ là ước lượng chệch của phương sai của ước lượng đúng $\hat{\theta}$. Thậm chí ngay cả khi X_1 và X_2 không tương quan thì nó vẫn là ước lượng chệch do:

$$E(D\hat{\alpha}_1) = D\hat{\theta}_1 + \frac{\theta_2^2 \sum x_{2t}^2}{\sum x_{1t}^2}; x_{1t} = X_{1t} - \overline{X}; x_{2t} = X_{2t} - \overline{X} \qquad (4.20)$$

6. Kết quả khoảng tin cậy thông thường và các quy tắc kiểm định không còn đúng nữa.

Ví dụ 4.6: Ta có bảng 4.4 biểu thị mối quan hệ giữa chi tiêu cho nhập khẩu (triệu USD) và thu nhập của một quốc gia trong 20 năm.

Bảng 4.4

Năm	Chi tiêu cho nhập khẩu	Thu nhập	Năm	Chi tiêu cho nhập khẩu	Thu nhập
1	35,7	1551,3	11	274,1	2167,4
2	144,6	1599,8	12	277,9	2212,6
3	150,9	1668,1	13	253,6	2214,3
4	166,2	1728,4	14	158,7	2248,6
5	190,7	1797,4	15	249,5	2261,5
6	218,2	1916,3	16	282,2	2331,9
7	211,8	1896,9	17	251,1	2469,8
8	187,9	1931,7	18	167,9	2542,8
9	299,9	2001,0	19	421,3	2640,9
10	259,4	2066,6	20	439,0	2686,3

Giả sử mô hình đúng biểu thị sự phụ thuộc giữa chi tiêu cho nhập khẩu (Y) với tiêu dùng (X_1) và biến xu thế (X_2) như sau :

$$y_t = \theta_0 + \theta_1 x_{1t} + \theta_2 x_{2t} + \xi_t \qquad (4.21)$$

với y_t, x_{1t}, x_{2t} là các giá trị tại quan sát thứ t (t = 1,2,..,20)

Mô hình như trong (4.21) cho ta biết có một biến khác là biến xu thế X_2 cũng ảnh hưởng tới chi tiêu cho nhập khẩu. Nó có thể là một biến: thời gian hoặc xu thế.

Nhưng thay cho ước lượng một mô hình đúng ta lại đi ước lượng mô hình sau :

$$y'_t = \alpha_0 + \alpha_1 x_{t1} + \gamma_t \qquad (4.22)$$

Dựa vào các giá trị có trong bảng 4.4, ta ước lượng được mô hình (4.22) như sau :

$$\hat{Y}' = -261{,}0914 + 0{,}2452312 X_1 \qquad (4.22')$$

Ước lượng cho độ lệch chuẩn :

$$\sqrt{\hat{D}(\hat{\alpha}_0)} = 31,3266; \ t_{\hat{\alpha}_0} = \frac{\hat{\alpha}_0}{\sqrt{\hat{D}(\hat{\alpha}_0)}} = -8,334$$

$$\sqrt{\hat{D}(\hat{\alpha}_0)} = 0,01476; \ t_{\hat{\alpha}_0} = \frac{\hat{\alpha}_1}{\sqrt{\hat{D}(\hat{\alpha}_1)}} = 6,616$$

Để phân tích kết quả chúng ta tiếp tục ước lượng hồi quy :

$$y_t = \theta_0 + \theta_1 x_{1t} + \theta_2 x_{2t} + \gamma_t \quad (4.23)$$

và hồi quy $x_{2t} = \beta_0 + \beta_1 x_{1t} + \varepsilon_t$ (4.24)

Dựa vào bảng giá trị 4.4 ta ước lượng được mô hình (4.23)

$$\hat{Y} = -859,9217 + 0,6499926X_1 - 23,19518X_2 \quad (4.23')$$

$$\sqrt{\hat{D}(\hat{\theta}_0)} = 11,9588; t_{\hat{\theta}_0} = \frac{\hat{\theta}_0}{\sqrt{\hat{D}(\hat{\theta}_0)}} = -7,681$$

$$\sqrt{\hat{D}(\hat{\theta}_1)} = 0,0745; \ t_{\hat{\theta}_1} = \frac{\hat{\theta}_1}{\sqrt{\hat{D}(\hat{\theta}_1)}} = 8,680$$

$$\sqrt{\hat{D}(\hat{\theta}_2)} = 4,2704; \ t_{\hat{\theta}_2} = \frac{\hat{\theta}_2}{\sqrt{\hat{D}(\hat{\theta}_2)}} = -5,432$$

Ước lượng cho mô hình (4.24)

$$\hat{X}_2 = -25,81701 + 0,0173212X_1 \quad (4.24')$$

$$\sqrt{\hat{D}(\hat{\beta}_0)} = 1,0757; \ t_{\hat{\theta}_0} = \frac{\hat{\beta}_0}{\sqrt{\hat{D}(\hat{\beta}_0)}} = -23,9999$$

$$\sqrt{\hat{D}(\hat{\beta}_1)} = 0,0173212; \ t_{\hat{\theta}_1} = \frac{\hat{\beta}_1}{\sqrt{\hat{D}(\hat{\beta}_1)}} = 34,177$$

+) Với mô hình hồi quy kiểm định sai (4.22) ta thấy rằng :

$$\frac{d\hat{Y}'}{dX_2} = 0,2453212$$

Khi thu nhập tăng lên 1triệu USD thì chi tiêu cho nhập khẩu tăng lên là 0,2453212 triệu USD. Trong khi đó mô hình **đúng (4.23)** cho thấy khi thu nhập

tăng lên 1 triệu USD thì chi tiêu cho nhập khẩu tăng lên hẳn 0,6499 do :

$$\frac{d\hat{Y}}{dX_2} = 0,6499$$

Trong trường hợp này mô hình chỉ định sai, ước lượng đúng lớn hơn ước lượng khi thiếu biến X_2. Bản chất của sự giảm xuống có thể thấy từ mô hình ước lượng hồi quy (4.24'). Ta thấy rằng $\hat{\beta}_1 = 0,0173212$; $\hat{\theta}_2 + \hat{\theta}_2\hat{\beta}_1 \approx 0,24523 \approx \hat{\alpha}_1$.

Như vậy việc bỏ sót biến X_2 không những đã bỏ qua ảnh hưởng của X_2 đối với chi tiêu cho nhập khẩu Y mà còn làm ảnh hưởng tới biến X_1. Vì biến X_1 phải gánh sự bỏ sót này (thấy được do hệ số của mô hình đúng giảm từ 0,649996 xuống 0,245321).

+) Hệ số chặn cũng bị chệch, nhưng ở đây ước lượng cho hệ số chặn với mô hình sai là –261,0914 lớn hơn ước lượng với mô hình đúng (-859,9217).

+) Phương sai của các sai số ước lượng được của hai mô hình cũng khác nhau.

+) Sai số tiêu chuẩn ở hai mô hình khác nhau.

Từ những nhận xét trên ta thấy rằng nếu ta định kiểm định giả thiết dựa trên mô hình sai thì những kết luận của chúng ta sẽ đáng ngờ, sẽ gây ra những hậu quả nghiêm trọng.

*) Đưa vào một số biến không thích hợp

Giả sử cần ước lượng mô hình đúng:

$$Y_t = \theta_0 + \theta_1 X_{1t} + \xi_t \tag{4.25}$$

Nhưng lại đem ước lượng mô hình

$$Y_t = \alpha_0 + \alpha_1 X_{1t} + \alpha_2 X_{2t} + \lambda_t \tag{4.26}$$

Như vậy sẽ mắc những sai lầm sau:

1. Các ước lượng bình phương nhỏ nhất cho hệ số hồi quy của mô hình (4.26) là những ước lượng không chệch và là ước lượng vững, nghĩa là:

$$E\hat{\alpha}_0 = \theta_0,\ \ E\hat{\alpha}_1 = \theta_1,\ \ E\hat{\alpha}_1 = 0$$

2. Ước lượng của σ^2 thu được từ mô hình hồi quy (4.26) là ước lượng vững.

3. Tuy nhiên các ước lượng thu được trong mô hình (4.26) là không hiệu quả. Các phương sai của chúng sẽ lớn hơn các phương sai của các ước lượng thu được trong (4.25). Do vậy nên khoảng tin cậy dựa trên các sai số tiêu chuẩn của mô hình (4.26) sẽ lớn hơn các khoảng tin cậy dựa trên các sai số tiêu chuẩn của các ước lượng từ mô hình đúng (4.25), dù cho khoảng tin cậy là chấp nhận được đối với kiểm định.

*) Dạng hàm không đúng

Việc ta dùng không đúng dạng hàm, sẽ đưa ra những ước lượng sai lệch, khoảng tin cậy để kiểm định không đúng và do đó sẽ đưa ra những kết luận sai.

3.7.3. Phát hiện những sai lầm chỉ định. Kiểm định sai lầm chỉ định

*) Phát hiện sự sai lầm của các biến không cần thiết

Giả sử ta có mô hình hồi quy sau:

$$Y_t = \theta_0 + \theta_1 X_{0t} + \theta_2 X_{2t} + \theta_3 X_{3t} + \theta_4 X_{4t} + \xi_t \tag{4.27}$$

Trong quá trình đưa ra mô hình, lý thuyết cho rằng tất cả các biến X_1, X_2, X_3, X_4 đều ảnh hưởng đến Y. Vậy nên ta phải giữ nguyên chúng trong mô hình mặc dù trong quá trình kiểm nghiệm mô hình ta thấy một biến nào đó không có trong mô hình. Ta có thể dùng các quy tắc kiểm định như ở trong mục 2.3 chương 3 của cuốn sách này.

*) Kiểm định các biến bị bỏ sót

Giả sử chúng ta kiểm định mô hình tuyến tính sau:

$$Y_t = \theta_0 + \theta_1 X_{1t} + \xi_t \tag{4.28}$$

Để kiểm định xem mô hình có bị thiếu một biến nào không ta phải ước lượng mô hình:

$$Y_t = \theta_0 + \theta_1 X_{1t} + \theta_2 Z_t + \xi_t \tag{4.29}$$

Nếu ta có số liệu sẵn về Z_t thì không sao, nhưng cái cần làm là kiểm định xem Z_t có bằng 0 hay không?

Trường hợp không có sẵn số liệu về Z_t ta sẽ sử dụng xấp xỉ $\hat{Z}$ đối với Z. Khi đó kiểm định thích hợp với biến bị bỏ sót là ước lượng mô hình:

$$Y_t = \theta_0 + \theta_1 X_{1t} + \theta_2 \hat{Z}_t + \lambda_t \tag{4.30}$$

và kiểm định giả thiết xem $\theta_2 = 0$ có bằng 0 hay không?

Người ta đưa ra quy tắc kiểm định d.Durbin-Watson để xem có mắc phải sai lầm bỏ sót biến hay không?

Tiêu chuẩn d. Durbin-Watson dùng để kiểm định có sự sai lầm định dạng hay không. Tuy nhiên không thể sử dụng trực tiếp mô hình này. Thủ tục này bao gồm:

1. Ước lượng mô hình ban đầu, chẳng hạn:

$$Y_t = \theta_0 + \theta_1 X_t + \xi_t$$

Tính các ước lượng cho sai số.

2. Nếu nghi ngờ biến Z bị bỏ sót, xếp các sai số theo thứ tự tăng dần của biến Z (nếu giá trị của Z có sẵn). Hoặc xếp các sai số theo thư tự tăng dần của biến độc lập nào đó.

3. Tính DW

$$DW = \frac{\sum_{t=2}^{n}(\hat{\xi}_t - \hat{\xi}_{t-1})^2}{\sum_{t=1}^{n}\hat{\xi}_t^2}$$

4. Kiểm định:

H_0: Dạng hàm là đúng.

H_1: Dạng hàm là sai.

Dựa vào thống kê DW và bảng phân phối Durbin-Watson để đưa ra kết luận.

§4. PHƯƠNG SAI CỦA SAI SỐ THAY ĐỔI

Khi nghiên cứu mô hình hồi quy tuyến tính cổ điển, chúng ta giả thiết: phương sai của các biến nhiễu là không đổi khi chuyển từ thí nghiệm này sang thí nghiệm khác và đều bằng σ^2, nghĩa là: $D\xi = \sigma^2$

Ngược lại với giả thiết chúng ta vẫn thường làm, tức là phương sai thay đổi khi tiến hành các thí nghiệm khác nhau.

4.1. Nguyên nhân hiện tượng phương sai của sai số thay đổi

Phương sai của sai số thay đổi do một số nguyên nhân sau:

- Do bản chất của các mối liên hệ kinh tế trong một hiện tượng kinh tế nào đó. Ngay trong bản thân của nó đã chứa đựng khả năng khi tiến hành các thí nghiệm để mô hình hoá cho hiện tượng kinh tế đó, sẽ làm cho phương sai của các sai số bị thay đổi. Chẳng hạn như khi xem xét mối quan hệ giữa thu nhập và tiết kiệm, thông thường thu nhập tăng lên thì mức độ biến động của tiết kiệm cũng tăng lên.

- Do kỹ thuật thu thập số liệu ngày càng được cải tiến, phương sai dường như giảm xuống.

- Phương sai của sai số thay đổi cũng xuất hiện khi có quan sát ngoại lai. Quan sát ngoại lai là các quan sát khác biệt rất nhiều với các quan sát khác trong mẫu. Việc đưa vào hay loại bỏ quan sát này ảnh hưởng rất lớn đến phân tích hồi quy.

- Mô hình đưa ra sai. Có thể bỏ mất biến hoặc thêm vào biến...

4.2. Ước lượng bình phương nhỏ nhất cho các hệ số hồi quy khi phương sai của sai số thay đổi

Ta xét mô hình hồi quy tuyến tính cổ điển dưới dạng ma trận (3.4) thoả mãn các giả thiết như đã nêu ở mục 2.1 trong chương 3, nhưng không thoả mãn điều kiện phương sai là không đổi.

Theo mệnh đề 3.1 trong chương 3 ta tính được ước lượng cho các hệ số hồi quy

$$\hat{\theta} = (X^T X)^{-1} X^T Y$$

Phương sai khi đó sẽ là:

$$D(\hat{\theta}) = D\left[(X^T X)^{-1} X^T Y\right]$$
$$= \left[(X^T X)^{-1} X^T\right]^2 D(Y) = B.D(X\theta + \xi)$$
$$= BD(\xi)$$

trong đó $B = \left[(X^T X)^{-1} X^T\right]^2$. Do phương sai của sai số là thay đổi nên $D(\hat{\theta}) = BD(\xi)$ cũng thay đổi theo. Vậy nên việc cần phải kiểm định các hệ số của mô hình và kiểm định lại mô hình.

4.3. Phát hiện ra phương sai của sai số thay đổi

Về mặt lý thuyết có thể dễ dàng phát hiện ra được phương sai của sai số thay đổi. Nhưng việc phát hiện ra vấn đề này trong thực tế không đơn giản do chúng ta chỉ biết được σ_i^2 khi đã biết được toàn tổng thể nghiên cứu. Chính vì thế người ta phải nghiên cứu để tìm ra phương pháp phát hiện ra phương sai của sai số là thay đổi.

Một số phương pháp để phát hiện phương sai của sai số thay đổi:

*) Bản chất của vấn đề nghiên cứu.

*) Xem xét đồ thị của sai số và biến phụ thuộc.

*) Kiểm định PARK.

*) Kiểm định GREJSER.

*) Kiểm định GOLDFELD – QUANDT.

*) Kiểm định BREUSCH –PAGAN – GODFREY (BPG).

*) Kiểm định trên biến phụ thuộc.

Trong đồ án này, trình bày những phương pháp sau:

4.3.1. Bản chất của vấn đề nghiên cứu

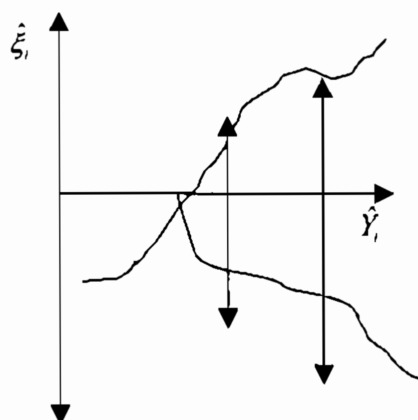
- Thông thường ngay trong bản chất của vấn đề nghiên cứu đã cho ta biết có thể xảy ra hiện tượng phương sai của sai số thay đổi hay không? Trên thực tế ta thấy rằng khi số liệu ở dạng chéo liên quan đến các nơi nghiên cứu khác nhau hay xảy ra hiện tượng phương sai của sai số là thay đổi.

- Chẳng hạn khi xem xét tình hình sản xuất của các doanh nghiệp sản xuất giày. Ta thấy rằng trong mẫu gồm những doanh nghiệp có quy mô khác nhau nên thường dẫn đến mô hình tổng quát đưa ra có phương sai của sai số là thay đổi.

4.3.2. Xem xét đồ thị của sai số và biến độc lập

Đồ thị của sai số với giá trị của biến độc lập nào đó, hoặc với giá trị ước lượng $\hat{Y}$ sẽ cho ta biết liệu phương sai của sai số có thay đổi hay không?

Nếu độ rộng của đồ thị tăng hoặc giảm khi $\hat{Y}_i$ tăng thì phương sai không phải là hằng số. Theo đồ thị hình 4.1 thì khả năng phương sai của sai số bị thay đổi là có thể xảy ra.



Hình 4.1

4.3.3. Kiểm định Park

- Park đã hình thức hoá phương pháp đồ thị rằng σ_t^2 là hàm số nào đó của biến Z_t :

$$\sigma_t^2 = \sigma^2 Z_t^b e^{\gamma_t} \qquad (4.31)$$

với Z_t có thể là :

+) Biến độc lập nếu đó là hàm hồi quy hai biến:

$\sigma_t^1 = \sigma^2 X_t^b e^{\gamma_t}$, với γ_t là sai số

+) Trường hợp hàm hồi quy bội thì có thể lấy σ_t^2 là một hàm của một biến độc lập nào đó, nhưng để tổng quát, ta lấy đó là một hàm phụ thuộc vào $\hat{Y}_t$, với $\hat{Y}_t$ là ước lượng của Y.

$$\sigma_t^2 = \sigma^2 Y_t^b e^{\gamma_t}$$

- Park đề nghị sử dụng $\hat{\xi}_i$ thay cho σ_t^2 và ước lượng hồi quy sau:

$$\ln \hat{\xi}_t^2 = \ln \sigma^2 + b \ln Z_t + \gamma_t \qquad (4.32)$$

Đặt $a = \ln \sigma^2$ thì (3.34) trở thành:

$$\ln \hat{\xi}_t^2 = a + b \ln Z_t + \gamma_t \qquad (4.33)$$

Như vậy để thực hiện kiểm định Park ta làm các bước sau:

1. Thực hiện ước lượng các hệ số cho mô hình hồi quy để tìm ra phương trình hồi quy thực nghiệm.

2. Thu được sai số sau đó đem bình phương và lấy ln.

3. Tuỳ từng loại mô hình hồi quy là hai biến hay hồi quy bội mà ta đem ước lượng mô hình (4.33).

4. Kiểm định giả thuyết $H_0 : a = 0$ nghĩa là không có hiện tượng phương sai của sai số thay đổi. Đối thuyết $H_1 : a \neq 0$. tức là phương sai của sai số bị thay đổi.

Ví dụ 4.7 : Xét tiếp ví dụ 3.1 trong chương 3. Theo ví dụ này mô hình đưa ra là mô hình hồi quy bội, vậy nên ta sẽ xét kiểm định Park với $Z_i = \hat{Y}_i$.

(4.33) lúc này trở thành:

$$\ln \hat{\xi}_i^2 = a + b \ln \hat{Y}_i + \gamma_i \ := \ Y = a + bX + \gamma_i \qquad (4.34)$$

Dựa vào $\hat{Y}_i ; \hat{\xi}_i$ đã có ta lập bảng 4.5 để tính :

Bảng 4.5

I	y_i	$\hat{y}_i$	$\hat{\xi}_i$	$\ln \hat{Y}_i$	$\ln \hat{\xi}_i^2$
1	127	124,9666	2,033	4,8280	1,4190
2	149	127,2659	1,734	4,9922	1,1009
3	106	108,4382	-2,438	4,686	1,7824
4	163	168,5537	-5,554	5,1272	3,4290
5	102	103,1741	-1,174	4,6364	0,3208
6	180	178,3238	1,676	5,1836	1,0328
7	161	161,5420	-0,542	5,085	-1,2298
8	128	129,4733	-1,473	4,8635	0,7746
9	139	131,979	7,021	4,8826	3,8978
10	144	147,0132	-3,013	4,9906	2,2059
11	159	154,0249	4,975	5,0371	3,2089
12	138	141,2437	-3,244	4,9505	2,3536

$$\hat{a} = \frac{\sum_{i=1}^{n}(\ln\hat{y}_i)^2\sum_{i=1}^{n}\ln\hat{\xi}_i^2 - \sum_{i=1}^{n}\ln\hat{y}_i\sum_{i=1}^{n}\ln\hat{y}_i\ln\hat{\xi}_i^2}{n\sum_{i=1}^{n}(\ln\hat{y}_i)^2 - (\sum_{i=1}^{n}\ln\hat{y}_i)^2}$$

$$= \frac{292{,}9824.20{,}3007 - 59{,}2629.100{,}4903}{12.292{,}9824 - 59{,}2629^2} \approx -2{,}055$$

$$\hat{b} = \frac{n\sum_{i=1}^{n}\ln\hat{\xi}_i^2\ln\hat{y}_i - \sum_{i=1}^{n}\ln\xi_i^2\sum_{i=1}^{n}\ln\hat{y}_i}{n\sum_{i=1}^{n}(\ln\hat{y}_i)^2 - (\sum_{i=1}^{n}\ln\hat{y}_i)^2}$$

$$= \frac{12.100{,}4903 - 20{,}3007.59{,}2629}{12.292{,}9824 - 59{,}2629^2} \approx 0{,}7597$$

Vậy phương trình hồi quy thực nghiệm như sau: $\hat{Y} = -2{,}055 + 0{,}7597X$

Ta sẽ kiểm định xem thực sự a có tham gia vào phương trình (4.34) hay không? Để tiện tính toán, ta lập bảng 4.6.

Bảng 4.6

i	y_i	$\hat{y}_i$	$\hat{\xi}_i$	$\ln\hat{Y}_i$	$\ln\hat{\xi}_i^2$	$\widehat{\ln\hat{\xi}_i^2}$	$\hat{\gamma}_i$
1	127	124,9666	2,033	4,8280	1,4190	1,6128	-0,1938
2	149	127,2659	1,734	4,9922	1,1009	1,7376	-0,6367
3	106	108,4382	-2,438	4,686	1,7824	1,5049	0,2775
4	163	168,5537	-5,554	5,1272	3,4290	1,8404	1,5886
5	102	103,1741	-1,174	4,6364	0,3208	1,4623	-1,1415
6	180	178,3238	1,676	5,1836	1,0328	1,883	-0,8502
7	161	161,5420	-0,542	5,085	-1,2298	1,8081	-3,0397
8	128	129,4733	-1,473	4,8635	0,7746	1,6400	-0,8654
9	139	131,979	7,021	4,8826	3,8978	1,6543	2,2435
10	144	147,0132	-3,013	4,9906	2,2059	1,7364	0,4695
11	159	154,0249	4,975	5,0371	3,2089	1,7717	1,4372
12	138	141,2437	-3,244	4,9505	2,3536	1,7059	0,6477

Áp dụng công thức (3.20) trong chương 3 ta có:

$$\hat{\sigma}'^2 = \frac{\hat{\gamma}^T\hat{\gamma}}{n-k-1} = \frac{\sum_{i=1}^{n}\hat{\gamma}_i^2}{n-k-1} = \frac{22,786}{10} = 2,2786$$

$$\hat{D}(\hat{a}) = \hat{\sigma}'^2 \frac{\sum_{i=1}^{n}(\ln\hat{y}_i)^2}{n\sum_{i=1}^{n}(\ln\hat{y}_i)^2 - (\sum_{i=1}^{n}\ln\hat{y}_i)^2} = 2,2786\frac{292,9824}{12.292,9824 - 59,2629} = 100,5529$$

$$=> \sqrt{\hat{D}(\hat{a})} = 13,437$$

$$\hat{D}(\hat{b}) = \frac{n}{n\sum_{i=1}^{n}(\ln\hat{y}_i)^2 - (\sum_{i=1}^{n}\ln\hat{y}_i)^2} = 2,2786\frac{12}{12.292,9824 - 59,2629} = 7,3951$$

$$=> \sqrt{\hat{D}(\hat{\theta}_1)} = 2,7194$$

$$=> t_{\hat{a}} = \frac{\hat{a}}{\sqrt{\hat{D}(\hat{a})}} = \frac{-2,055}{13,437} = -0,1529\ ; => t_{\hat{b}} = \frac{\hat{b}}{\sqrt{\hat{D}(\hat{b})}} = \frac{0,7597}{2,7194} = 0,2794$$

Theo mệnh đề 3.4 ở chương 3 nếu chọn α = 0,05 thì $t_{n-k-1}\left(\frac{\alpha}{2(k+1)}\right)$ là phân vị trên mức $\frac{\alpha}{2(k+1)}$ của phân bố Student với n-k-1 bậc tự do.Tra bảng Student $t_{10}\left(\frac{0,05}{2(1+1)}\right) \approx 2,76$ => $|t_{\hat{a}}| < t_{10}\left(\frac{0,05}{2(1+1)}\right)$; $|t_{\hat{b}}| < t_{10}\left(\frac{0,05}{2(1+1)}\right)$ $\Rightarrow$ Chấp nhận H_0 $\Rightarrow$ Không có hiện tượng phương sai của sai số bị thay đổi.

4.3.4. Kiểm định Goldfeld-Quandt

Ta nói σ_i^2 có liên quan dương với một trong các biến của mô hình hồi quy nếu:

$$\sigma_i^2 = \sigma^2 X_i^2 \qquad (4.35)$$

Với X_i là một biến độc lập của mô hình hồi quy tuyến tính cổ điển

Khi phương sai của sai số thay đổi σ_i^2 có liên quan dương với một trong các biến giải thích trong hàm hồi quy thì ta có thể áp dụng kiểm định này.

Thủ tục kiểm định của Goldfeld-Quandt gồm các bước như sau:

Bước 1: Sắp xếp các quan sát theo thứ tự tăng dần của một biến X_i

Bước 2: Bỏ c quan sát ở giữa theo cách sau:

+) $c = 4$ nếu cỡ mẫu khoảng $n = 30$,

+) $c = 10$ nếu cỡ mẫu khoảng $n = 60$,

và chia số quan sát còn lại thành hai nhóm, trong đó mỗi nhóm có $(n - c)/2$ quan sát.

Bước 3: Sử dụng phương pháp BPCT để ước lượng các tham số của các hàm hồi quy đối với $(n - c)/2$ quan sát đầu và cuối. Sau khi ước lượng, ta sẽ thu được các phần dư tương ứng với hai mô hình hồi quy đó.

Gọi tổng bình phương các sai số của mô hình có giá trị biến độc lập nhỏ hơn và lớn hơn lần lượt là: RSS1; RSS2. Bậc tự do tương ứng của chúng đều là $(n - c - 2k)/2$.

Bước 4: Tính thống kê $F = \dfrac{RSS2}{RSS1}$

Nếu điều kiện biến nhiễu của mô hình hồi quy tuyến tính ban đầu có phân phối chuẩn và phương sai của nó là không đổi. Khi đó F có phân phối Fisher với bậc của tử và mẫu đều bằng $(n - c - 2k)/2$

Ta đưa ra quy tắc kiểm định như sau:

Giả thuyết H_0: Phương sai của sai số là không đổi σ_i^2 = const.

Đối thuyết H_1: Phương sai của sai số là thay đổi.

4.3.5. Kiểm định Breusch – Pagan - Godfrey (BPG)

Phần 4.3.4 ta đã xét kiểm định Gold - Quandt không chỉ phụ thuộc vào số các giá trị bị bỏ đi mà còn phụ thuộc vào biến độc lập nào được lấy ra để sắp xếp các giá trị cho kiểm định. Kiểm định BGP sẽ khắc phục được nhược điểm này.

Ta sẽ xét mô hình hồi quy tuyến tính cổ điển sau:

$$Y = \theta_0 + \theta_1 X_1 + \ldots + \theta_k X_k + \xi \qquad (4.36)$$

Giả sử ở quan sát thứ i ($i = 1, 2, .., n$) được mô tả như hàm của biến phi ngẫu nhiên Z_i, Z_i ở đây là các biến X_i (là một số biến hay tất cả các biến) có ảnh hưởng đến σ_i^2. Hàm của σ_i^2 có dạng :

$$\sigma_i^2 = f(z_{1i}, z_{2i}, \ldots, z_{mi})$$

Hàm $f(z)$ có dạng hàm tuyến tính hoặc là hàm loga. Đặc biệt ta giả thiết

$$\sigma_i^2 = \alpha_0 + \alpha_1 Z_{1i} + \alpha_2 Z_{2i} ... + \alpha_m Z_{mi} \tag{4.37}$$

Từ (4.37) ta thấy rằng nếu $\alpha_1 = \alpha_2 = ... = \alpha_m$ thì $\sigma_i^2 = \alpha_0$ (hằng số). Nhờ vào đặc điểm này mà có thể đưa ra kiểm định

H_0: $\alpha_1 = \alpha_2 = ... = \alpha_m$ tức là phương sai có sai số là không đổi.

Thủ tục kiểm định như sau:

Bước 1: Ước lượng mô hình hồi quy ban đầu bằng phương pháp BPCT để thu được các phần sai số

Bước 2: Tính $\tilde{\sigma}^2 = \frac{\sum_{i=1}^{n} \hat{\xi}_i^2}{n}$

Bước 3: Xây dựng biến $p_i = \frac{\hat{\xi}_i^2}{\tilde{\sigma}^2}$

Bước 3: Hồi quy p_i theo các biến Z_i dưới dạng:

$$p_i = \alpha_0 + \alpha_1 Z_{1i} + \alpha_2 Z_{2i} + ... + \alpha_m Z_{mi} + \gamma_i \tag{4.38}$$

trong đó γ_i là số hạng ngẫu nhiên của hồi quy (4.38).

Bước 4: Sau khi ước lượng cho mô hình (4.38) ta thu

ESS = $\sum_{i=1}^{n} (\hat{p}_i - \overline{p})^2$ trong đó $\hat{p}_i$ là ước lượng của p_i. Sau đó xác định

$$\tau = \frac{1}{2} \text{ESS} \tag{4.39}$$

Kiểm định này giả thiết các nhiễu (sai số) có phân phối chuẩn và khi kích thước mẫu n tăng lên vô hạn thì $\tau \approx \chi^2_{(m-1)}$. Nên nếu trong quá trình tính toán ta tính được τ mà thấy rằng $\tau > \chi^2_{(m-1)}$ thì ta bác bỏ H_0, tức là bác bỏ giả thuyết phương sai của sai số là không đổi.

Ví dụ 4.8. Cho bộ số liệu sau đây:

X_1: Kinh nghiệm công tác (số năm có việc làm đầy đủ).

X_2: số năm được đào tạo.

Y: Thu nhập trung bình trên một giờ.

Bộ số liệu cho dưới dạng bảng 4.7.

Bảng 4.7

i	X_1	X_2	Y	i	X_1	X_2	Y	i	X_1	X_2	Y
1	0	6,0	4,71	18	18,0	0	5,06	35	34,0	2,0	9,98
2	1,0	3,0	3,60	19	19,0	6,0	13,69	36	37,0	6,0	27,73
3	2,0	0	4,37	20	21,0	0	8,01	37	37,0	0	5,06
4	2,0	4,0	4,64	21	21,0	2,0	17,13	38	37,0	1,0	4,36
5	3,0	1,0	3,27	22	23,0	1,0	7,75	39	38,0	7,0	23,96
6	5,0	0	4,26	23	24,0	0	6,2	40	38,0	4,0	30,77
7	6,0	70	6,14	24	24,0	5,0	17,72	41	39,0	0	20,68
8	7,0	5,0	6,74	25	24,0	3,0	8,80	42	40,0	2,0	50,90
9	8,0	0	6,11	26	25,0	2,0	12,8	43	42,0	3,0	3,96
10	8,0	2,0	5,53	27	25,0	0	5,2	44	42,0	0	7,58
11	8,0	6,0	5,53	28	27,0	4,0	8,12	45	43,0	4,0	6,18
12	10,0	1,0	5,36	29	28,0	7,0	17,54	46	44,0	3,0	43,25
13	11,0	7,0	8,73	30	28,0	4,0	22,52	47	44,0	1,0	32,04
14	13,0	0	5,85	31	30,0	3,0	5,47	48	45,0	0	3,35
15	15,0	0	6,88	32	31,0	1,0	13,67	49	45,0	2,0	18,35
16	15,0	2,0	7,17	33	32,0	0	4,84	50	46,0	0	4,95
17	15,0	7,0	10,8	34	34,0	5,0	38,52				

Ước lượng bằng phương pháp BPCT ta thu được phương trình thực nghiệm như sau: $\hat{Y} = -1,5971 + 0,40979X_1 + 1,4609X_2$ (*)

Ta đưa ra kiểm định chung sau đây với α = 0,05

H_0: Phương sai của sai số là không đổi.

H_1: Phương sai của sai số là thay đổi.

+) Kiểm định Goldfeld-Quandt

Ta thấy rằng biến X_1 được sắp xếp theo thứ tự tăng dần. Ở đây có tất cả 50 quan sát, ta sẽ bỏ đi c = 10 quan sát ở giữa từ mẫu thứ 21 đến mẫu 30. Từ đó ta có hai mẫu có kích thước bằng nhau có 20 quan sát. Ta đem ước lượng bằng phương pháp BPCT để thu được các phương trình ước lượng tương ứng.

Mẫu 1: Từ quan sát thứ 1 đến 20

$$\hat{Y}^1 = 2,1528 + 0,30129X_1 + 0,47597X_2; \sum_{i=1}^{20} \hat{\xi}_i^2 = 26,8225$$

Mẫu 2: Từ quan sát thứ 30 đến 50

$$\hat{Y}^2 = -7,7849 + 0,47968X_1 + 3,1414X_2; \quad \sum_{i=30}^{50} \hat{\xi}_i^2 = 3363,0$$

Tính thống kê

$$F = \frac{\sum_{i=30}^{50} \hat{\xi}_i^2 /((n-c-2k)/2)}{\sum_{i=1}^{20} \hat{\xi}_i^2 ((n-c-2k)/2)} = \frac{3363,0}{26,8224} \approx 125,38$$

Tra bảng Fisher với α = 0,05. Bậc của tử và mẫu đều bằng $(n - c - 2k)/2$ = 17. Ta có $F_{(17;\ 17,\ 0,05)} \approx 2,3 \Rightarrow F > F_{(17;\ 17;\ 0,05)}$. Bác bỏ H_0 và đưa ra kết luận: Phương sai của sai số là thay đổi.

+) Kiểm định BPG

Từ mô hình (*) đã ước lượng, ta tính được các sai số và tính được

$$\sum_{i=1}^{50} \hat{\xi}_i^2 = 3975,8 \Rightarrow \tilde{\sigma}^2 = \frac{\sum_{i=1}^{50} \hat{\xi}_i^2}{n} = \frac{3975,8}{50} = 79,516\,.$$

Sau đó tính các p_i theo công thức sau: $p_i = \frac{\hat{\xi}_i^2}{\tilde{\sigma}^2}$. Trên cơ sở đó có mô hình

$$p_i = \alpha_0 + \alpha x_{i1} + \alpha_2 x_{i2} + \gamma_i \,(**)$$

Ước lượng mô hình (**) bằng phương pháp BPCT ta thu được phương trình ước lượng như sau: $p_i = 1,6151 - 6,1645X_i - 0,032X_2$

$$\text{ESS} = \sum_{i=1}^{n} (\hat{p}_i - \overline{p})^2 = 48,986; \; \tau = \frac{1}{2}\text{ESS} = 24,493$$

Tra bảng χ với mức ý nghĩa α = 0,05 và bậc tự do là k − 1 = 2 có $\chi^2_{0,05,2} \approx 5,99 \Rightarrow \tau > \chi^2_{0,05;2} \Rightarrow$ Bác bỏ H_0, tức là ở đây có hiện tượng phương sai của sai số thay đổi.

CHƯƠNG 5

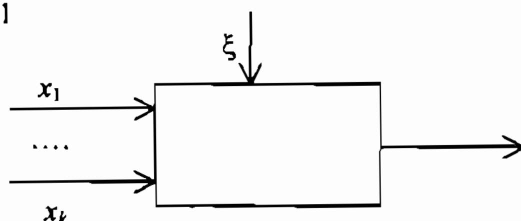
QUY HOẠCH TRỰC GIAO

Dựa trên phương pháp BPCT, chỉ khác là chủ động bố trí ma trận X, do đó các kết quả của phương pháp BPCT đều áp dụng được cho Quy hoạch trực giao (QHTG). Dựa vào các định nghĩa của phương pháp BPCT, ta suy ra các tính chất của QHTG, đồng thời sẽ thấy rõ ưu điểm của phương pháp này so với phương pháp BPCT.

§1. QUY HOẠCH TRỰC GIAO VÀ TÍNH CHẤT

1.1. Mở đầu

Xét mô hình 5.1



Hình 5.1

Làm n thí nghiệm để đo Y, ta được bảng 5.1 như sau:

Bảng 5.1

N⁰	X_1	X_2	...	X_k	Y
1	x_{11}	x_{12}	...	x_{1k}	y_1
2	x_{21}	x_{22}	...	x_{2k}	y_2
...	...	...	...	...	...
n	x_{n2}	x_{n2}	...	x_{nk}	y_n

Giả thiết:

$$Y = \theta_0 + \theta_1 X_1 + \theta_2 X_2 + \dots + \theta_k X_k + \xi \qquad (5.1)$$

$$\xi \sim N(0, \sigma^2) \qquad (5.2)$$

$$\tilde{Y} = EY = \theta_0 + \theta_1 X_1 + \theta_2 X_2 + \dots + \theta_k X_k \qquad (5.3)$$

trong đó θ_j ; $j = 0, \ldots, k$ là các hệ số chưa biêt.

Bằng phương pháp BPCT, ta ước lượng được:

$$\hat{\theta} = (X^T X)^{-1} X^T Y \tag{5.4}$$

trong đó:

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} \quad \hat{\theta} = \begin{bmatrix} \theta_0 \\ \theta_0 \\ \vdots \\ \theta_k \end{bmatrix}$$

Tính được các hệ số hồi quy, thay vào phương trình (5.3) ta được

$$\hat{Y} = \hat{\theta}_0 + \hat{\theta}_1 X_1 + \ldots + \hat{\theta}_k X_k \tag{5.5}$$

Phương trình (5.5) chính là phương trình hồi quy thực nghiệm.

Bây giờ ta xét vấn đề liệu có thể bố trí các thí nghiệm sao cho:

+ Số thí nghiệm ít.

+ Tính toán gọn.

+ Bảo đảm mức độ chính xác.

1.2. Định nghĩa QHTG

QHTG là quy hoạch bố trí các thí nghiệm sao cho ma trận:

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}$$

Có tính chất: $\sum\limits_{m=1}^{n} x_{im} x_{ij} = 0$ (5.6)

Trong đó i: chỉ số của thí nghiệm.

m, j: chỉ số biến $m, j = 0, 1, \ldots, k$.

Khi $m = 0$ thì $x_{i0} = 1$ với $\forall i = 1, 2, \ldots, n$, vậy nên từ (5.6) ta suy ra

$$\sum_{i=1}^{n} x_{ij} = 0; \forall i \neq 0 \tag{5.7}$$

Ý nghĩa (5.6): Tích vô hướng của hai vector cột bất kỳ bằng 0. Đó chính là tính chất trực giao.

Ý nghĩa (5.7): Tổng các phần tử trong một cột bất kỳ (trừ cột 0) đều bằng 0

1.3. Tính chất của QHTG

Tính chất 1: Công thức tính $\hat{\theta}_j$ đơn giản:

Theo mệnh đề 3.1 trong chương 3 thì : $\hat{\theta} = (X^T X)^{-1} X^T Y$

$$X^T X = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_{11} & x_{21} & \dots & x_{n1} \\ \dots & \dots & \dots & \dots \\ x_{1m} & x_{2m} & \dots & x_{nm} \\ \dots & \dots & \dots & \dots \\ x_{1k} & x_{2k} & \dots & x_{nk} \end{pmatrix} \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1j} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & xj & \dots & x_{2k} \\ 1 & \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nj} & \dots & x_{nk} \end{pmatrix}$$

$$= C = (C_{mj})_{(k+1)(k+1)}$$

$$C_{mj} = \sum_{i=1}^{n} x_{im} x_{ij} = 0; \quad j \neq m$$

Ký hiệu $C_j^2 = \sum_{i=1}^{n} x_{ij}^2 \Rightarrow X^T X = \begin{pmatrix} C_0^2 & & & 0 \\ & C_1^2 & & \\ & & \dots & \\ 0 & & & C_k^2 \end{pmatrix} = C$

$$(X^T X)^{-1} = C^{-1} = \begin{pmatrix} \frac{1}{C_0^2} & & & 0 \\ & \frac{1}{C_1^2} & & \\ & & \dots & \\ 0 & & & \frac{1}{C_k^2} \end{pmatrix}$$

$$X^T Y = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_{11} & x_{21} & \dots & x_{n1} \\ \dots & \dots & \dots & \dots \\ x_{1k} & x_{2k} & \dots & x_{nk} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_{i1} y_i \\ \dots \\ \sum_{i=1}^{n} x_{ik} y_i \end{pmatrix}$$

$$\hat{\theta} = (X^T X)^{-1}(X^T Y) = \begin{pmatrix} \frac{1}{C_0^2} & & & 0 \\ & \frac{1}{C_1^2} & & \\ & & \cdots & \\ 0 & & & \frac{1}{C_k^2} \end{pmatrix} \begin{pmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_{i1} y_i \\ \cdots \\ \sum_{i=1}^{n} x_{ik} y_i \end{pmatrix} = \begin{pmatrix} \frac{1}{C_0^2} \sum_{i=1}^{n} y_i \\ \frac{1}{C_1^2} \sum_{i=1}^{n} y_i \\ \cdots \\ \frac{1}{C_k^2} \sum_{i=1}^{n} y_i \end{pmatrix}$$

Vậy

$$\left.\begin{aligned} \hat{\theta}_0 &= \frac{1}{C_0^2} \sum_{i=1}^{n} y_i = \frac{1}{n} \sum_{i=1}^{n} y_i \\ \hat{\theta}_j &= \frac{1}{C_j^2} \sum_{i=1}^{n} x_{ij} y_i, \qquad j = 1, 2, \ldots, k \end{aligned}\right\} \tag{5.8}$$

Tính chất 2

+ $E(\hat{\theta}_j) = \theta_j; j = 0, 1, \ldots, k$ (5.9)

$\Rightarrow$ $\hat{\theta}_j$ là ước lượng không chệch của θ_j

+) $D(\hat{\theta}_j) = \sigma^2 \left\{(X^T X)\right\}_{jj}^{-1} = \dfrac{\sigma^2}{C_j^2}; j = 0, 1, \ldots, k$ (5.10)

$\Rightarrow$ Phương sai của $\hat{\theta}_j$ càng bé nếu σ^2 bé. Và ngược lại.

+) $\text{cov}(\hat{\theta}_j, \hat{\theta}_m) = 0;\ j \neq m$ (5.11)

$\Rightarrow$ $\hat{\theta}_j, \hat{\theta}_m$ là không tương quan với nhau.

Tính chất này có lợi khi tăng bậc của đa thức

Tính chất 3:

+) $E(\hat{y}_i) = \tilde{y}_i$ (5.12)

$\Rightarrow$ $\hat{y}_i$ là ước lượng trúng của $\tilde{y}_i$

+) $D(\hat{y}_i) = \sigma^2 \left\{(XC^{-1}X^T)\right\}_{ii} = \sigma^2 \sum_{j=0}^{k} \dfrac{x_{ij}^2}{C_j^2}$ (5.13)

$\Rightarrow$ Ta có thể tính được phương sai của $\hat{y}_i$

+) $\text{cov}\left(\hat{y}_p, \hat{y}_q\right) = \sigma^2 \left\{(XC^{-1}X^T)\right\}_{pq} = \sigma^2 \sum_{j=0}^{k} \dfrac{x_{pj} x_{qj}}{C_j^2}$ (5.14)

⇒ Ta sẽ thấy được tính tương quan của $\hat{y}_p, \hat{y}_q$

$$XC^{-1}X^T = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1j} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2j} & \dots & x_{2k} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nj} & \dots & x_{nk} \end{pmatrix} \begin{pmatrix} \frac{1}{C_0^2} & & & 0 \\ & \frac{1}{C_1^2} & & \\ & & \dots & \\ 0 & & & \frac{1}{C_k^2} \end{pmatrix} \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_{11} & x_{21} & \dots & x_{n1} \\ \dots & \dots & \dots & \dots \\ x_{1k} & x_{2k} & \dots & x_{nk} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{1}{C_0^2} & \frac{x_{11}}{C_1^2} & \frac{x_{12}}{C_2^2} & \dots & \frac{x_{1k}}{C_k^2} \\ \frac{1}{C_0^2} & \frac{x_{21}}{C_1^2} & \frac{x_{22}}{C_2^2} & \dots & \frac{x_{2k}}{C_k^2} \\ \frac{1}{C_0^2} & \frac{x_{n1}}{C_1^2} & \frac{x_{n2}}{C_2^2} & \dots & \frac{x_{nk}}{C_k^2} \end{pmatrix} \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_{11} & x_{21} & \dots & x_{n1} \\ \dots & \dots & \dots & \dots \\ x_{1k} & x_{2k} & \dots & x_{nk} \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{j=0}^{k} \frac{x_{1j}^2}{C_j^2} & \sum_{j=0}^{k} \frac{x_{1j}x_{2j}}{C_j^2} & \dots & \sum_{j=0}^{k} \frac{x_{1j}x_{nj}}{C_j^2} \\ \sum_{j=0}^{k} \frac{x_{2j}x_{1j}}{C_j^2} & \sum_{j=0}^{k} \frac{x_{2j}^2}{C_j^2} & \dots & \sum_{j=0}^{k} \frac{x_{2j}x_{nj}}{C_j^2} \\ \dots & \dots & \dots & \dots \\ \sum_{j=0}^{k} \frac{x_{nj}x_{1j}}{C_j^2} & \sum_{j=0}^{k} \frac{x_{nj}x_{2j}}{C_j^2} & \dots & \sum_{j=0}^{k} \frac{x_{nj}^2}{C_j^2} \end{pmatrix}$$

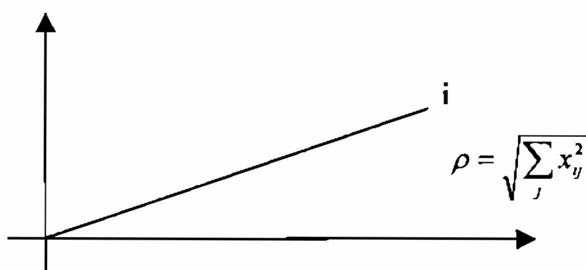
Nhận xét: Nếu C_j = const thì $C_j^2 = k$

$$D(\hat{y}_i) = \frac{\sigma^2}{k} \sum_{j=0}^{k} x_{ij}^2, \mathrm{cov}(y_p, y_q) = \frac{\sigma^2}{k} \sum_{j=0}^{k} x_{pj} x_{qj}$$

$$D(\hat{y}_i) = \frac{\sigma^2}{k} \left(x_{i0}^2 + \sum_{j=1}^{k} x_{ij}^2 \right) = \frac{\sigma^2}{k} (1 + \rho_i^2)$$

$\rho_i^2 = x_{i1}^2 + x_{i2}^2 + \dots + x_{ik}^2$ là bình phương khoảng cách từ gốc toạ độ đến điểm thí nghiệm i được minh hoạ trên hình 5.2.

Vậy nếu C_j = const thì phương sai của i, j chỉ phụ thuộc vào khoảng cách từ gốc toạ độ đến điểm thứ i. Người ta nghiên cứu thí nghiệm để $C_j^2 = const$.



Hình 5.2

§2. QUY HOẠCH TRỰC GIAO CẤP MỘT

2.1. Định nghĩa

QHTG cấp một là một cách bố trí thí nghiệm sao cho quy hoạch là QHTG và có thêm tính chất:

$$C_j^2 = \sum_{i=1}^{n} x_{ij}^2 = n; \qquad j = 0,1,...,k \tag{5.15}$$

Tức là tổng bình phương các phần tử của một cột đúng bằng số thí nghiệm

2.2. Tính chất

Từ các tính chất của QHTG ta suy ra các tính chất của QHTG cấp 1 sau:

Tính chất 1:

$$\left.\begin{aligned} \hat{\theta}_j &= \frac{1}{n}\sum_{i=1}^{n} x_{ij} y_i \\ \hat{\theta}_0 &= \frac{1}{n}\sum_{i=1}^{n} y_i \end{aligned}\right\} \tag{5.16}$$

Tính chất 2:

$$\left.\begin{aligned} E(\hat{\theta}_j) &= \theta_j \\ D\hat{\theta}_j &= \frac{1}{n}\sigma^2 \\ \operatorname{cov}(\hat{\theta}_j, \hat{\theta}_m) &= 0 \end{aligned}\right\} \tag{5.17}$$

Ta thấy rằng: số thí nghiệm càng lớn thì phương sai càng nhỏ.

Tính chất 3:

$$E\hat{y}_i = \tilde{y}_i$$

$$D(\hat{y}_i) = \frac{\sigma^2}{b}\sum_{j=0}^{k} x_{ij}^2 = \frac{\sigma^2}{n}\left(1 + \sum_{j=1}^{k} x_{ij}^2\right) = \frac{\sigma^2}{n}\left(1 + d_i^2\right); \quad i = 1, 2, \ldots, n$$

Trong đó: d_i là khoảng cách từ gốc toạ độ đến điểm thí nghiệm thứ i. Khi quay về gốc toạ độ thì phương sai này không đổi, là tính chất quay được.

Ý nghĩa: y_i phân tán đều với $\forall i$

2.3. Ma trận X của QHTG cấp một

- Khi $k = 2$, ta có mô hình có 3 biến : $\tilde{Y} = \theta_0 + \theta_1 X_1 + \theta_2 X_2$

$$X = \begin{pmatrix} 1 & -1 & -1 \\ 1 & +1 & -1 \\ 1 & -1 & +1 \\ 1 & +1 & +1 \end{pmatrix} \qquad n = 2^2 = 4$$

Ở đây để tránh nhầm lẫn, ta viết ± thay cho ±1.

- Khi $k = 3$; ta có mô hình 4 biến như sau:

$$\tilde{Y} = \theta_0 + \theta_1 X_1 + \theta_2 X_2 + \theta_3 X_3$$

$$X = \begin{pmatrix} X_0 & X_1 & X_2 & X_3 \\ + & - & - & - \\ + & + & - & - \\ + & - & + & - \\ + & + & + & - \\ + & - & - & + \\ + & + & - & + \\ + & - & + & + \\ + & + & + & + \end{pmatrix} \begin{matrix} \\ 1 \\ X_1 \\ X_2 \\ X_1X_2 \\ X_3 \\ X_1X_3 \\ X_2X_3 \\ X_1X_2X_3 \end{matrix} \qquad n = 2^3 = 8$$

- Khi $k = 4$: mô hình 5 biến như sau:

$$\tilde{Y} = \theta_0 + \theta_1 X_1 + \theta_2 X_2 + \theta_3 X_3 + \theta_4 X_4$$

$$X = \left(\begin{array}{ccccc}
X_0 & X_1 & X_2 & X_3 & X_4 \\
+ & - & - & - & - \\
+ & + & - & - & - \\
+ & - & + & - & - \\
+ & + & + & - & - \\
+ & - & - & + & - \\
+ & + & - & + & - \\
+ & - & + & + & - \\
+ & + & + & + & - \\
+ & - & - & - & + \\
+ & + & - & - & + \\
+ & - & + & - & + \\
+ & - & - & + & + \\
+ & + & + & - & + \\
+ & + & - & + & + \\
+ & - & + & + & + \\
+ & + & + & + & +
\end{array}\right)
\begin{array}{c}
\\
1 \\
X_1 \\
X_2 \\
X_1X_2 \\
X_3 \\
X_1X_3 \\
X_2X_3 \\
X_1X_2X_3 \\
X_4 \\
X_1X_4 \\
X_2X_4 \\
X_3X_4 \\
X_1X_2X_4 \\
X_1X_3X_4 \\
X_2X_3X_4 \\
X_1X_2X_3X_4
\end{array}$$

Với số thí nghiệm $n = 2^4 = 16$.

2.4. Mã hoá các biến (đổi biến)

Theo như trên thì các phần tử của ma trận X là +1; −1. Nhưng nói chung, khoảng biến thiên của các biến mà ta nghiên cứu khác [−1; +1]. Vậy làm thế nào để bố trí thí nghiệm sao cho X trực giao.

Gọi biến thực tế là Z_j, $j = 1, 2, \ldots, k$: $\underline{Z}_j \le Z_j \le \bar{Z}_j$ ta thu được

$$Z_j^0 = \frac{\bar{Z}_j + \underline{Z}_j}{2}; \quad \Delta Z_j = \frac{\bar{Z}_j - \underline{Z}_j}{2}; \quad x_j = \frac{Z_j - Z_j^0}{\Delta Z_j}$$

$$Z_j = \underline{Z}_j \Leftrightarrow x_j = -1$$

$$Z_j = Z_j^0 \Leftrightarrow x_j = 0$$

$$Z_j = \bar{Z}_j \Leftrightarrow x_j = 1$$

Và bảng 5.2 ma trận thí nghiệm Z

Bảng 5.2

N^0	X_0 X_1 X_2 ... X_k	Z_1 Z_2 Z_3 ... Z_k	Y

2.5. Kiểm định các kết quả

Việc kiểm định các giả thiết thống kê hoàn toàn giống như đã trình bày trong chương 3. Quá trình kiểm định bao gồm:

2.5.1. Kiểm định $D\xi = \sigma^2$

Vì không biết σ^2 nên ta dùng phương sai tái sinh để ước lượng σ^2. Nếu không có điều kiện làm nhiều thí nghiệm tại mỗi điểm ta có thể tính s_{ts}^2 như sau:

Làm n_0 thí nghiệm ở tâm tức là $x_1 = x_2 = ... = 0$, do đó được $y_0^1, y_0^2, ..., y_0^{n_0}$

$$s_{ts}^2 = \frac{1}{n_0 - 1}\sum_{i=1}^{n_0}(y_0^i - \bar{y}_0)$$

$$\bar{y}_0 = \frac{1}{n_0}\sum_{i=1}^{n_0} y_0^i$$

Bậc tự do là $n_0 - 1$.

2.5.2. Kiểm tra các hệ số hồi quy có bằng không hay không

Chú ý rằng khi có $\theta_j = 0$, nghĩa là x_j không thực sự ảnh hưởng trực tiếp đến y thì x_j có thể loại bỏ.

Trong trường hợp chung, phải làm lại với mô hình mới (không có x_j tham gia). Nhưng trong QHTG ta đã biết $\text{cov}(\hat{\theta}_j, \hat{\theta}_m) = 0$, tức là ở đây $\hat{\theta}_j, \hat{\theta}_m$ không tương quan nên không phải làm lại thí nghiệm và không phải tính các $\hat{\theta}_j \neq 0$ đã có.

2.5.3. Kiểm định sự phù hợp của phương trình hồi quy

Ở đây có thể dùng phương sai tái sinh s_{ts}^2 bằng cách làm m thí nghiệm, nhưng cũng có thể chỉ làm n_0 thí nghiệm ở tâm.

Ví dụ 5.1: Tìm quan hệ giữa Y và Z_1, Z_2, Z_3

$$0,008 \le Z_1 \le 0,1$$

$$0,03 \le Z_2 \le 0,3$$

$$19 \le Z_3 \le 84$$

Ta có bảng 5.3

Bảng 5.3

N^0	X_0	X_1	X_2	X_3	Z_1	Z_2	Z_3	Y
1	+	-	-	-	0,008	0,03	19	13,9
2	+	+	-	-	0,1	0,03	19	18,5
3	+	-	+	-	0,008	0,3	19	2
4	+	+	+	-	0,1	0,3	19	3
5	+	-	-	+	0,008	0,03	84	16
6	+	+	-	+	0,1	0,03	84	18,5
7	+	-	+	+	0,008	0,3	84	9
8	+	+	+	+	0,1	0,3	84	12

Giả sử mô hình tuyến tính mẫu $\tilde{Y} = \theta_0 + \theta_1 X_1 + \theta_2 X_2 + \theta_3 X_3$

*) Tính các hệ số $\hat{\theta}_j; j = 0,1,2,3$ trong mô hình thực nghiệm

$$\hat{Y} = \hat{\theta}_0 + \hat{\theta}_1 X_1 + \hat{\theta}_2 X_2 + \hat{\theta}_3 X_3$$

$$\hat{\theta}_0 = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{8}(13,9 + 18,5 + 2 + 3 + 16 + 18,5 + 9 + 12) = 11,613$$

$$\hat{\theta}_1 = \frac{1}{n}\sum_{i=1}^{n} x_{i1} y_i = 1,388$$

$$\hat{\theta}_2 = \frac{1}{n}\sum_{i=1}^{n} x_{i2} y_i = -5,112$$

$$\hat{\theta}_3 = \frac{1}{n}\sum_{i=1}^{n} x_{i3} y_i = 2,262$$

Vậy ta có mô hình hồi quy thực nghiệm:

$$\hat{Y} = 11,613 + 1,388 X_1 - 5,112 X_2 + 2,262 X_3$$

*) Kiểm định $\theta_j = 0$ với điều kiện đã làm 3 thí nghiệm ở tâm là:

$$y_0^1 = 11,1 \qquad ,y_0^2 = 11,3 \qquad ,y_0^3 = 10,8$$

$$\Rightarrow \bar{y}_0 = \frac{1}{3}(11,1 + 11,3 + 10,8) = 11,066$$

$$\Rightarrow s_{ts}^2 = \frac{1}{3-1}\left[(11,1 - 11,066)^2 + (11,3 - 11,066)^2 + (10,8 - 11,066)^2\right]$$

$$= \frac{11,067}{2} = 0,063$$

$$\Rightarrow s_{\hat{\theta}_j}^2 = s_{ts}^2 \{C^{-1}\}_{jj}, s_{\hat{\theta}_j} = \frac{s_{ts}}{\sqrt{n}} = 0,089$$

Tính $t_{\hat{\theta}_j} = \dfrac{\hat{\theta}_j}{s_{\hat{\theta}j}} \Rightarrow$

$$t_{\hat{\theta}_0} = \frac{11,613}{0,089} = 130,513 \qquad ; t_{\hat{\theta}_1} = \frac{1,388}{0,089} = 15,594$$

$$t_{\hat{\theta}_2} = \frac{-5,112}{0,089} = -57,460 \qquad ; t_{\hat{\theta}_3} = \frac{2,262}{0,089} = 25,428$$

Chọn α = 0,05 bậc tự do là $n_0 - 1 = 2$. Tra bảng Student ta được $t_\alpha = 2,92$. Vì $|t_{\hat{\theta}_j}| > t_\alpha$ nên bác bỏ giả thiết H_0 và kết luận các hệ số đều có nghĩa.

⇒ Phương trình hồi quy thực nghiệm có dạng:

$$\hat{Y} = 11,613 + 1,388X_1 - 5,112X_2 + 2,262X_3$$

*) Kiểm định sự phù hợp của mô hình:

Sau khi xây dựng được mô hình, ta đi tính các phương sai dư:

$$s_{du}^2 = \frac{1}{n-(k+1)}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \frac{26,78}{8-4} = 6,695$$

$$\hat{F} = \frac{s_{du}^2}{s_{t_s}^2} = \frac{6,695}{0,063} = 106,27$$

Chọn $\alpha = 0,05$, bậc của tử sẽ là $n - (k + 1) = 4$; bậc của mẫu là $n_0 - 1 = 2$. Tra bảng Fisher ta có được $F_{(4;2;0,05)} = 19,2 \Rightarrow \hat{F} > F_{(4;2,0,05)} = 19,2$. Vậy nên bác bỏ giả thiết H_0, tức là không chấp nhận mô hình. Điều này sẽ được giải quyết trong phần sau.

2.6. Một số mô hình mở rộng từ QHTG cấp một

2.6.1. Mô hình 1

QHTG cấp một có thể xác định được các hệ số của phương trình hồi quy dạng $\tilde{Y} = \sum_{j=0}^{k} \theta_j X_j$

Nếu kiểm định không phù hợp, ta có thể giả thiết mô hình có dạng bậc hai không hoàn chỉnh.

$$\tilde{Y} = \theta_0 + \theta_1 X_1 + \theta_2 X_2 + \theta_{12} X_1 X_2$$

Trường hợp tổng quát: $\tilde{Y} = \sum_{j=0}^{k} \theta_j X_j + \sum_{i,j=0}^{k} \theta_{ij} X_i X_j$

Khi đó bằng cách đặt $X_3 = X_1X_2$ ta được: $\tilde{Y} = \theta_0 + \theta_1 X_1 + \theta_2 X_2 + \theta_{12} X_3$. Ta sẽ được $\tilde{Y} = \theta_0 + \theta_1 X_1 + \theta_2 X_2 + \theta_{12} X_3$.

Khi đó ma trận quy hoạch có dạng như bảng 5.4

Bảng 5.4

n	X_0	X_1	X_2	$X_3 = X_1X_2$
1	+	–	–	+
2	+	+	–	–
3	+	–	+	–
4	+	+	–	+

2.6.2. Mô hình 2

Mô hình dạng: $\tilde{Y} = a_0 + a_1 X^1 + a_2 X^2 ... + a_n X^n$

Với mô hình này bằng cách đặt biến mới $X_1 = X, X_2 = X^2, ..., X_n = X^n$ ta sẽ có mô hình tuyến tính như sau:

$$\tilde{Y} = a_0 + a_1 X_1 + a_2 X_2 ... + a_n X_n$$

Dùng QHTG cấp một xác định các hệ số. Nếu không được ta phải dùng phương pháp BPCT.

2.6.3. Một số mô hình khác

Với những mô hình $y = f(x)$ cũng có thể biến đổi trở thành tuyến tính bằng cách đặt các ẩn phụ hoặc đổi biến. Do vậy cũng có thể dùng QHTG cấp một.

Việc làm đó không phải chỉ với mô hình hàm một biến mà ngay cả đối với hàm nhiều biến miễn là điều kiện cho phép, nghĩa là sau khi đổi biến thì tính chất trực giao của ma trận X vẫn được đảm bảo.

Nếu trường hợp bậc 2 không hoàn chỉnh và kiểm định không phù hợp nữa thì phải giả thiết mô hình bậc 2 là hoàn chỉnh.

$$\tilde{Y} = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_{12} x_1 x_2 + \theta_{11} x^2{}_1 + \theta_{22} x_2^2$$

$$\tilde{Y} = \sum_{j=0}^{k} \theta_j x_j + \sum_{i<j}^{k} \theta_{ij} x_i x_j + \sum_{j=1}^{k} \theta_{ij} x_j^2$$

Trong trường hợp này, nếu cũng biến đổi như trong mô hình 1 ở mục 2.6.1 thì ma trận X không còn tính chất trực giao nữa. Ta phải tiến hành quy hoạch thí nghiệm kiểu khác. Ở trong đồ án này, quy hoạch khác này chính là QHTG cấp hai sẽ được xét ở phần sau.

§3. QUY HOẠCH TRỰC GIAO CẤP HAI

3.1. Khái niệm về QHTG cấp hai

QHTG cấp hai là phương pháp xây dựng quy trình thí nghiệm và xử lý số liệu trên một số tiêu chí giống QHTG cấp một, nhưng ở đây ta sẽ nhận được mô hình hồi quy dạng đa thức bậc hai đủ mô tả sự phụ thuộc của hàm Y vào các thông số ảnh hưởng $X_1, X_2, \ldots, X_k$

Xét mô hình bậc hai đầy đủ

$$Y = \theta_0 + \sum_{j=1}^{k} \theta_j X_j + \sum_{i<j}^{k} \theta_{ij} X_i X_j + \sum_{j=1}^{k} \theta_{ij} X_j^2 + \xi$$

Với $\xi \sim N(0,\sigma^2)$; $D\xi \sim \sigma^2$

Đặt $\tilde{Y} = EY = \theta_0 + \sum_{j=1}^{k} \theta_j X_j + \sum_{i<j}^{k} \theta_{ij} X_i X_j + \sum_{j=1}^{k} \theta_{ij} X_j^2 + \xi$

Vấn đề đặt ra ở đây phải quy hoạch thí nghiệm thế nào để được ước lượng mô hình thống kê gần đúng tốt nhất.

Người ta đề nghị một cách bố trí thí nghiệm như sau gọi là GHTG cấp hai. Gồm 3 loại thí nghiệm:

- Loại 1 (phần cơ sở): Gồm $n_1 = 2k$ thí nghiệm hoàn toàn giống như QHTG cấp một. Yêu cầu đặt ra là khi xây dựng phần cơ sở phải đảm bảo tính được tất

cả các hệ số hồi quy tuyến tính θ_j và tương tác cặp đôi θ_{ij}. Tác động của chúng không bị lẫn vào nhau.

- Loại 2 (phần tâm): Gồm n_0 ($n_0 \geq 1$) thí nghiệm ở tâm miền quy hoạch. Tại đó giá trị mã của các thông số bằng không: $X_1 = 0$, $X_2 = 0$, ..., $X_k = 0$ ứng với $Z_1^0, Z_2^0, ..., Z_k^0$.

- Loại 3: Gồm $n_k = 2k$ thí nghiệm bố trí trên các trục toạ độ, cách gốc toạ độ một đoạn $\alpha > 0$ sao cho ma trận X trực giao, tức là lấy $X_j = \pm\alpha$ với $j = 0, 1, 2, ..., k$.

$\Rightarrow$ Tổng số thí nghiệm $n = 2^k + n_0 + 2k$. Để đỡ nhầm lẫn với số biến ta sẽ ký hiệu chỉ số thí nghiệm là u, $u = 1, 2, ..., n$.

Trong QHTG cấp hai, ta sẽ xét các vấn đề:

- Xây dựng ma trận X

- Tìm các hệ số mô hình

- Kiểm định các kết quả

3.2. Các vấn đề cần làm trong việc xây dựng QHTG cấp hai

3.2.1. Xây dựng ma trận X

Ta xét mô hình bậc $k = 2$ (khi $k > 2$ cách làm tương tự)

$$\tilde{Y} = \theta_0 + \theta_1 X_1 + \theta_2 X_2 + \theta_{12} X_1 X_2 + \theta_{11} X_1^2 + \theta_{22} X_2^2$$

Với $\xi \sim (0, \sigma^2)$; $D\xi = \sigma^2$

Lúc này: $n_1 = 2^2 = 4$; $n_0 = 1$; $n_k = 2k$. Ma trận X có dạng:

$$X = \begin{pmatrix} X_0 & X_1 & X_2 & X_1X_2 & X_1^2 & X_2^2 \\ + & - & - & + & + & + \\ + & + & - & - & + & + \\ + & - & + & - & + & + \\ + & + & + & + & + & + \\ + & 0 & 0 & 0 & 0 & 0 \\ + & \alpha & 0 & 0 & \alpha^2 & 0 \\ + & -\alpha & 0 & 0 & \alpha^2 & 0 \\ + & 0 & \alpha & 0 & 0 & \alpha^2 \\ +, & 0 & -\alpha & 0 & 0 & \alpha^2 \end{pmatrix} \begin{matrix} \\ \left.\begin{matrix} \\ \\ \\ \\ \end{matrix}\right\} n_1 = 2^2 \\ \left.\begin{matrix} \\ \end{matrix}\right\} n_0 = 1 \\ \left.\begin{matrix} \\ \\ \\ \\ \end{matrix}\right\} n_k = 4 \end{matrix}$$

Nếu ta chọn α và hai cột X_1^2, X_2^2 không khéo thì ma trận không có tính chất trực giao vì:

$$\sum_{u=1}^{n} x_{u_0} x_{u_j}^2 \neq 0 \qquad j = 1,2; \sum_{u=1}^{n} x_{u_i}^2 x_{u_j} \neq 0 \qquad i \neq j$$

Để giải quyết khó khăn đó ta làm biến đổi sau:

$X_j^2 \rightarrow X_j'$ với $X_j' = X_j^2 - \lambda$; $j = 0, 1, 2, \ldots, k$

Ma trận X sẽ như sau:

$$X = \begin{pmatrix} X_0 & X_1 & X_2 & X_1X_2 & X_1^2 - \lambda & X_2^2 - \lambda \\ + & - & - & + & 1-\lambda & 1-\lambda \\ + & + & - & - & 1-\lambda & 1-\lambda \\ + & - & + & - & 1-\lambda & 1-\lambda \\ + & + & + & + & 1-\lambda & 1-\lambda \\ + & 0 & 0 & 0 & -\lambda & -\lambda \\ + & \alpha & 0 & 0 & \alpha^2-\lambda & -\lambda \\ + & -\alpha & 0 & 0 & \alpha^2-\lambda & -\lambda \\ + & 0 & \alpha & 0 & -\lambda & \alpha^2-\lambda \\ + & 0 & -\alpha & 0 & -\lambda & \alpha^2-\lambda \end{pmatrix}$$

Khi quan sát ma trận X ta thấy phải lập điều kiện trực giao cho từng cặp một. Do tính chất đặc biệt của ma trận X nên thực chất chỉ cần có hai điều kiện dành cho trường hợp tổng quát:

+) Điều kiện k cột cuối của X phải trực giao với các cột đầu:

$$2^k + 2\alpha^2 - (2^k + 2k + n_0)\lambda = 0$$

$$\Rightarrow \lambda = \frac{2^k + 2\alpha^2}{2^k + 2k + n_0} = \frac{2^k + 2\alpha^2}{n} \qquad (5.18)$$

$$\Rightarrow X_j' = X_j^2 - \frac{1}{n}(2^k + 2\alpha^2) \text{ với mọi k}$$

+) Điều kiện là k cột cuối trực giao với nhau:

$$\Rightarrow \alpha = \sqrt{\sqrt{n.2^{k-2}} - 2^{k-1}} \text{ cho mọi } k \qquad (5.19)$$

Bây giờ thử lại với $k = 2$

$$\alpha = \sqrt{\sqrt{9.2^{2-2}} - 2^{2-1}} = \sqrt{\sqrt{9} - 2} = \sqrt{3-2} = 1$$

$$X'_j : X'_j = X_j^2 - \frac{1}{9}(2^2 + 2\times 1) = X_j^2 - \frac{2}{3}$$

Xét X'_1 thấy X_1^2 bằng nhau trong 4 dòng đầu:

$u = 1, 2, 3, 4 \qquad x'_{u_1} = 1 - \frac{2}{3} = \frac{1}{3}$

$u = 5 \qquad x'_{u_1} = 0 - \frac{2}{3} = -\frac{2}{3}$

$u = 6, 7 \qquad x'_{u_1} = 1 - \frac{2}{3} = \frac{1}{3}$

$u = 8, 9 \qquad x'_{u_1} = 0 - \frac{2}{3} = -\frac{2}{3}$

Tương tự với X_2^2. Bây giờ ta sẽ thu được ma trận X như sau:

$$X = \begin{pmatrix} X_0 & X_1 & X_2 & X_1X_2 & X'_1 & X'_2 \\ + & - & - & + & 1/3 & 1/3 \\ + & + & - & - & 1/3 & 1/3 \\ + & - & + & - & 1/3 & 1/3 \\ + & + & + & + & 1/3 & 1/3 \\ + & 0 & 0 & 0 & -2/3 & -2/3 \\ + & 1 & 0 & 0 & 1/3 & -2/3 \\ + & -1 & 0 & 0 & 1/3 & -2/3 \\ + & 0 & 1 & 0 & -2/3 & 1/3 \\ + & 0 & -1 & 0 & -2/3 & 1/3 \end{pmatrix}$$

Ma trận này có tính trực giao.

Bảng bố trí thí nghiệm được nêu trong bảng 5.5.

Bảng 5.5

N^0	X_0	X_1	X_2	X_1X_2	X'_1	X'_2	Z_1	Z_2
1 .. N								

Ta vẫn dùng công thức mã hoá như sau: $X_j = \dfrac{Z_j - Z_j^0}{\Delta Z_j}$

- Khi $j=\overline{1,2^k}$ kết quả như cũ, $j \in n_0$ thì $Z_j - Z_j^0$

- Khi $j=\overline{1,2k}$ đối với $X_j = \pm - 1$ cũng có tương tự $X_j = \pm - 1$ hay $Z_j = \underline{Z}_j$

- Khi $X_j = 0$ thì $Z_j = Z_j^0$

3.2.2. Tính toán và kiểm định

Dùng công thức cũ nhưng phải chú ý đến sự thay đổi trong ma trận X. Kết hợp với điều kiện (5.18), (5.19) ta dẫn đến sự thay đổi trong cách tính:

$$C = X^T X = \begin{pmatrix} N & & & & & & & & & 0 \\ & 2^k + 2\alpha^2 & & & & & & & & \\ & & \dots & & & & & & & \\ & & & 2^k + 2\alpha^2 & & & & & & \\ & & & & 2^k & & & & & \\ & & & & & \dots & & & & \\ & & & & & & 2^k & & & \\ & & & & & & & 2\alpha^4 & & \\ & & & & & & & & \dots & \\ 0 & & & & & & & & & 2\alpha^4 \end{pmatrix}$$

$$\Rightarrow C^{-1} = (X^T X)^{-1} = \begin{pmatrix} \frac{1}{N} & & & & & & & & & 0 \\ & \frac{1}{2^k + 2\alpha^2} & & & & & & & & \\ & & \dots & & & & & & & \\ & & & \frac{1}{2^k + 2\alpha^2} & & & & & & \\ & & & & \frac{1}{2^k} & & & & & \\ & & & & & \dots & & & & \\ & & & & & & \frac{1}{2^k} & & & \\ & & & & & & & \frac{1}{2\alpha^4} & & \\ & & & & & & & & \dots & \\ 0 & & & & & & & & & \frac{1}{2\alpha^4} \end{pmatrix}$$

Dễ dàng ta nhận thấy:

$$\sum_{u=1}^{n} x_{uj}^2 = 2^k + 2\alpha^2 \qquad \sum_{u=1}^{n}(x_{ui}x_{uj})^2 = 2^k \qquad \sum_{u=1}^{n}(x'_{uj})^2 = 2\alpha^4$$

+) Tiến hành như trong (5.16), ta được các công thức ước lượng cho các hệ số hồi quy:

$$\hat{\theta}_0 = \frac{1}{n}\sum_{u=1}^{n} y_u ; \hat{\theta}_j = \frac{\sum_{u=1}^{n} x_{uj} y_u}{\sum_{u=1}^{n} x_{uj}^2}$$ với số hạng bậc nhất $j = \overline{1,k}$

$$\hat{\theta}_{ij} = \frac{\sum_{u=1}^{n}(x_{ui}x_{uj})y_u}{\sum_{u=1}^{n}(x_{ui}x_{uj})^2}$$ với số hạng chéo.

$$\hat{\theta}_{jj} = \frac{\sum_{u=1}^{n} x'_{ui} y_u}{\sum_{u=1}^{n}(x'_{uj})^2}$$ với số hạng bậc hai.

+) Để kiểm định $D\hat{\xi} = \sigma^2$ vẫn dùng phương sai tái sinh theo các nghiệm ở tâm

+) Để kiểm định các hệ số hồi quy có bằng 0 hay không, thì phải tính các phương sai $D(\hat{\theta}_j) = \sigma^2 \left\{C^{-1}\right\}_{ij}$

$$s_{\hat{\theta}_0}^2 = \frac{s_{ts}^2}{n} \qquad s_{\hat{\theta}_j}^2 = \frac{s_{ts}^2}{\sum_{u=1}^{n} x_{uj}^2}$$

$$s_{\hat{\theta}_{ij}}^2 = \frac{s_{ts}^2}{\sum_{u=1}^{n}(x_{ui}x_{uj})^2} \qquad s_{\hat{\theta}_{jj}}^2 = \frac{s_{ts}^2}{\sum_{u=1}^{n}(x'_{ui})^2}$$

+) Để kiểm định sự phù hợp của mô hình tiến hành so sánh hai phương sai s_{ts}^2 và s_{du}^2 như cũ với s_{du}^2 được tính như trong QHTG cấp một.

$$s_{du}^2 = \frac{1}{n-(k'+1)}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

trong đó k' chính là số biến còn lại của mô hình thực nghiệm đã được kiểm định $\hat{Y}$.

3.3. Ví dụ 5.2:

Tiếp tục tìm mối quan hệ của Y và Z_1, Z_2, Z_3 ở ví dụ 5.1

$0,008 \le Z_1 \le 0,1$

$0,03 \le Z_2 \le 0,3$

$19 \le Z_3 \le 84$

Trong ví dụ trên đã tìm ra mô hình tuyến tính bậc 1 cho bài toán, nhưng sau khi kiểm định thấy mô hình không phù hợp. Vì vậy phải tiếp tục dùng QHTG cấp hai để xây dựng mô hình phù hợp hơn.

+) Mã hoá và lập ma trận thí nghiệm

Ta giả sử mô hình tuân theo dạng sau:

$$\tilde{Y} = \theta_0 + \sum_{j=1}^{3}\theta_j X_j + \sum_{\substack{j=1\\ i<j}}^{3}\theta_{ij} X_i X_j + \sum_{j=1}^{3}\theta_{jj} X_{jj}^2$$

Để lập ma trận thực nghiệm, trước tiên ta phải chuyển Z_j sang X_j bằng cách mã hoá theo công thức

$$X_j = \frac{Z_j - Z_j^0}{\Delta Z_j}$$

Sau đó lập bảng thí nghiệm đầy đủ. Số thí nghiệm với $k = 3$

$n = n_0 + n_1 + n_k = 2^3 + 4 + 2.3 = 18$, với $n_0 = 4$ là số thí nghiệm tại tâm

$$\alpha = \sqrt{\sqrt{n.2^{k-2}} - 2^{k-1}} = \sqrt{\sqrt{18.2^1} - 2^2} = 1,414$$

$$X_j' = X_j^2 - \frac{1}{n}(2^k + 2\alpha^2) = X_j^2 - \frac{2}{3}$$

Giá trị của Z_j; X_0, X_j ; $j = 1, 2, 3$ và Y được cho trong bảng 5.6 sau:

Bảng 5.6

N^0	X_0	X_1	X_2	X_3	Z_1	Z_2	Z_3	Y
1	+	-	-	-	0,008	0,03	19	13,9
2	+	+	-	-	0,1	0,03	19	18,5
3	+	-	+	-	0,008	0,3	19	2
4	+	+	+	-	0,1	0,3	19	3
5	+	-	-	+	0,008	0,03	84	16
6	+	+	-	+	0,1	0,03	84	18,5
7	+	-	+	+	0,008	0,3	84	9
8	+	+	+	+	0,1	0,3	84	12

Sau khi đổi biến ta tính toán và lập bảng 5.7

Bảng 5.7

N^0	X_0	X_1	X_2	X_3	X_1X_2	X_1X_3	X_2X_3	X'_1	X'_2	X'_3	Y
1	+	–	–	–	+	+	+	1/3	1/3	1/3	13.9
2	+	+	–	–	–	–	+	1/3	1/3	1/3	18.5
3	+	–	+	–	–	+	–	1/3	1/3	1/3	2
4	+	+	+	–	+	–	–	1/3	1/3	1/3	3
5	+	–	–	+	+	–	–	1/3	1/3	1/3	16
6	+	+	–	+	–	+	–	1/3	1/3	1/3	18,5
7	+	–	+	+	–	–	+	1/3	1/3	1/3	9
8	+	+	+	+	+	+	+	1/3	1/3	1/3	12
9	+	0	0	0	0	0	0	–2/3	–2/3	–2/3	10,1
10	+	0	0	0	0	0	0	–2/3	–2/3	–2/3	11,2
11	+	0	0	0	0	0	0	–2/3	–2/3	–2/3	9,9
12	+	0	0	0	0	0	0	–2/3	–2/3	–2/3	12,3
13	+	α	0	0	0	0	0	1,33	–2/3	–2/3	15
14	+	–α	0	0	0	0	0	1,33	–2/3	–2/3	8
15	+	0	α	0	0	0	0	–2/3	1,33	–2/3	7.5
16	+	0	–α	0	0	0	0	–2/3	1,33	–2/3	15,8
17	+	0	0	α	0	0	0	–2/3	-2/3	1,33	11,5
18	+	0	0	–α	0	0	0	–2/3	-2/3	1,33	5

+) Ta đi ước lượng cho các hệ số hồi quy

Dựa vào các công thức ở mục 3.3 ta tính toán để ước lượng các hệ số hồi quy. Giá trị tính toán của chúng như sau:

$\hat{\theta}_0 = 11,067$ $\quad\hat{\theta}_1 = 1,75$ $\quad\hat{\theta}_2 = -4,386$ $\quad\hat{\theta}_3 = 2,274$

$\hat{\theta}_{12} = -0,388$ $\quad\hat{\theta}_{13} = -0,013$ $\quad\hat{\theta}_{23} = 1,737$

$\hat{\theta}_{11} = 0,763$ $\quad\hat{\theta}_{22} = 0,838$ $\quad\hat{\theta}_{33} = -0,862$

+) Kiểm định $D\xi = \sigma^2$

Dùng 4 thí nghiệm ở tâm, với các thí nghiệm ở tâm đã cho như trong bảng hình 5.8. Cũng trong bảng này ta sẽ tính giá trị trung bình cho các giá trị ở tâm.

Bảng 5.8

n_0	y_0^t	Giá trị	$\bar{y}_0$
1	y_0^1	10,1	10,875
2	y_0^2	11,2	
3	y_0^3	9,9	
4	y_0^4	12,3	

Phương sai tái sinh

$$s_{ts}^2 = \frac{1}{n_0 - 1}\sum_{t=1}^{n_0}(y_0^t - \bar{y}_0)^2 = 1,229$$

+) Kiểm định các hệ số

$$s_{\hat{\theta}_0}^2 = \frac{s_{ts}^2}{n} = \frac{1,229}{18} = 0,068 \Rightarrow s_{\hat{\theta}_0} = 0,261$$

$$s_{\hat{\theta}_j}^2 = \frac{s_{ts}^2}{\sum_{u=1}^{n} x_{uj}^2} = \frac{1,229}{12} = 0,1024 \Rightarrow s_{\hat{\theta}_j} = 0,32$$

$$s_{\hat{\theta}_{ij}}^2 = \frac{s_{ts}^2}{\sum_{u=1}^{n}(x_{ui}x_{uj})^2} = \frac{1,229}{8} = 0,153 \Rightarrow s_{\hat{\theta}_{ij}} = 0,392$$

$$s^2_{\hat{\theta}_{jj}} = \frac{s^2_{ts}}{\sum_{u=1}^{n}(x'_{uj})^2} = \frac{1,229}{8} = 0,153 \Rightarrow s_{\hat{\theta}_{jj}} = 0,392$$

Các thống kê tính toán như sau:

$t_{\hat{\theta}_0} = 42,349$ $t_{\hat{\theta}_1} = 5,468$ $t_{\hat{\theta}_2} = -13,706$ $t_{\hat{\theta}_3} = 7,106$

$t_{\hat{\theta}_{12}} = -0,989$ $t_{\hat{\theta}_{13}} = -0,032$ $t_{\hat{\theta}_{23}} = 4,431$

$t_{\hat{\theta}_{11}} = 1,946$ $t_{\hat{\theta}_{22}} = 2,137$ $t_{\hat{\theta}_{33}} = -2,2$

Với mức ý nghĩa $\alpha = 0,05$, bậc tự do $n_0 - 1 = 3$. Tra bảng phân vị Student ta có $t_\alpha = 2,353$. Ta thấy $t_{\hat{\theta}_{12}}, t_{\hat{\theta}_{13}}, t_{\hat{\theta}_{11}}, t_{\hat{\theta}_{22}}, t_{\hat{\theta}_{33}}$ đều nhỏ hơn $t_\alpha = 2,353$. Vậy các hệ số của các biến sau trong mô hình bằng không là X_1X_2; X_1X_3; X_1^2; X_2^2; X_3^2.

Phương trình hồi quy thực nghiệm là:

$\hat{Y} = 11,067 + 1,75X_1 - 4,386X_2 + 2,274X_3 + 1,737X_2X_3$

+) Kiểm tra sự phù hợp của mô hình:

Ta đi tính phương sai dư

$$s^2_{du} = \frac{1}{n-(k'+1)}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \frac{1}{13}\sum_{i=1}^{18}(y_i - \hat{y}_i)^2 = \frac{42,252}{13} = 3,25$$

Xét tỷ số:

$$\hat{F} = \frac{s^2_{du}}{s^2_{ts}} = \frac{3,25}{1,229} = 2,644$$

Chọn $\alpha = 0,05$. Tra bảng phân vị Fisher với bậc của tử là 13, bậc của mẫu là 3. Ta có $F_{13,3,0,05} = 8,75 \Rightarrow \hat{F} < F_{13;3,0,05} \Rightarrow$ Chấp nhận mô hình.

Chuyển biến Z_j với $Z_j = X_j.\Delta Z_j + Z^0_j$

$\hat{Y} = 14,135 + 38,043Z_1 - 52,887Z_2 + 0,005Z_3 + 0,396Z_2Z_3$

CHƯƠNG 6

ÁP DỤNG MÔ HÌNH HỒI QUY VÀO PHƯƠNG PHÁP NGOẠI SUY ĐỂ DỰ BÁO

§1. GIỚI THIỆU KHÁI QUÁT VỀ DỰ BÁO KINH TẾ

1.1. Sự ra đời tất yếu của dự báo

Thuật ngữ "dự báo" đã được biết đến từ lâu, nói lên thuộc tính không thể thiếu của bộ não con người đó là sự phản ánh vượt trước. Sự ra đời của khoa học dự báo xuất phát từ những đòi hỏi thực tế, thể hiện ở các nội dung:

- Do chịu ảnh hưởng của cách mạng khoa học kỹ thuật;
- Do yêu cầu nâng cao công tác hoạch định;
- Do yêu cầu mở rộng hợp tác quốc tế.

Ngày nay, nhu cầu dự báo đã trở nên hết sức cần thiết ở mọi lĩnh vực, đặc biệt những dự báo kinh tế đã và đang đóng góp hữu hiệu cho việc quản lý kinh tế.

1.2. Các đặc điểm chung của các loại dự báo

Hiện nay có rất nhiều các loại dự báo khác nhau đang được sử dụng. Chúng khác nhau về nhiều khía cạnh, tuy nhiên có một vài đặc điểm nào đó là chung cho tất cả và việc nhận ra chúng là quan trọng. Dưới đây, ta nêu ra các đặc điểm chung đó:

1.2.1. Các kỹ thuật dự báo thường thừa nhận sự giống nhau dưới các hệ thống nhân quả đã tồn tại trong quá khứ sẽ tiếp tục tồn tại trong tương lai.

1.2.2. Các dự báo ít khi hoàn hảo, các kết quả thực tế thường khác so với các giá trị dự báo. Không ai có thể dự báo một cách chính xác được một số lớn các nhân tố liên quan thường xuyên tác động đến quá trình dự báo và sự có mặt của yếu tố ngẫu nhiên ngăn cản một dự báo hoàn hảo.

1.2.3. Các dự báo của một nhóm các mục thường có xu hướng chính xác hơn các dự báo của một mục đơn lẻ bởi sai số dự báo trong các mục của một nhóm có thể triệt tiêu nhau do có các giá trị âm, dương.

1.2.4. Độ chính xác của dự báo giảm đi khi khoảng thời gian bao trùm bởi các dự báo tăng. Nói chung, các dự báo ngắn hạn thường có yếu tố không chắc chắn ít hơn so với dự báo dài hạn.

1.3. Sơ đồ dự báo và các nguyên tắc dự báo

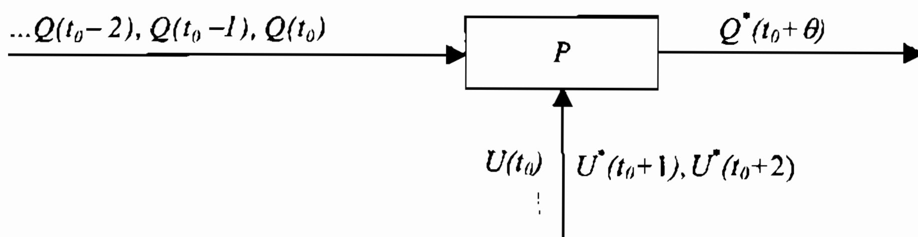
1.3.1. Sơ đồ dự báo

Giả sử có hệ thống S với trạng thái ở thời điểm t là:

$$Q(t) = (Q_1(t), Q_2(t), \ldots, Q_m(t)).$$

Nội dung của dự báo là căn cứ vào những thông tin về quá khứ và hiện tại của bản thân hệ thống đó, phối hợp với những thông tin của môi trường ngoài để đưa ra những phán đoán về tương lai của hệ thống.

Dựa vào đó mà ta có sơ đồ dự báo



Hình 6.1. Sơ đồ dự báo.

trong đó: $U(t_0)$ - thông tin hiện tại của môi trường;

P - toán tử dự báo;

$Q^*(t_0+\theta)$ - giá trị dự báo về trạng thái của hệ thống S tại thời điểm tương lai $t_0 + \theta > 0$ (tầm xa dự báo).

Công thức dự báo:

$$Q^*(t_0+\theta) = P\{Q(t_0), Q(t_0 - 1), \ldots, U(t_0), U^*(t_0+1)\} \quad (6.1)$$

Thông thường có thể đưa ra một loạt giá trị Q^* khác nhau ứng với mỗi độ tin cậy khác nhau. Đây là phương pháp dự báo nhiều phương án. Trong không gian Q xây dựng một chuẩn và $\delta = \left|Q^*(t_0+\theta) - Q(t_0+\theta)\right|$ là sai số của dự báo và cần tìm tiệm cận trên của dự báo ε với $\delta \le \varepsilon$.

Dung sai của dự báo: là sai số chấp nhận được ứng với mỗi vấn đề dự báo và tầm xa của dự báo nhất định.

1.3.2. Các nguyên tắc dự báo

a) Nguyên tắc liên hệ biện chứng:

Nguyên tắc này yêu cầu khi tiến hành dự báo một đối tượng kinh tế phải xem xét đến tính hệ thống và những nhân tố ảnh hưởng cùng vận động đồng thời.

Khi dự báo về tổng sản lượng quốc gia phải đồng thời phân tích dựa trên các mối quan hệ khác như: mức độ tiến bộ khoa học kỹ thuật, lượng lao động xã hội, đầu tư và tiết kiệm...

b) Nguyên tắc thừa kế lịch sử

Nguyên tắc này yêu cầu khi tiến hành dự báo cần phải nghiên cứu sâu sắc quá trình vận động của đối tượng trong quá khứ và hiện tại, tạo ra cơ sở thực nghiệm để tiên đoán và đánh giá tác động xu hướng trong tương lai.

c) Nguyên tắc đặc thù về bản chất đối tượng dự báo:

Nguyên tắc này đòi hỏi phải tính đến những nét đặc thù về bản chất của đối tượng cần dự báo, xuất phát từ những nét đặc thù này sẽ tạo cho chúng ta những giới hạn nhất định về xu thế phát triển trong tương lai.

d) Nguyên tắc mô tả tối ưu đối tượng dự báo

Nguyên tắc này yêu cầu: phải mô tả đối tượng dự báo bằng những mô hình tối ưu thông qua phương pháp mô tả hình thức và phi hình thức, đảm bảo giải quyết được nhiệm vụ dự báo với chi phí ít nhất, phải mô tả dự báo thông qua một lượng với số biến tối thiểu và đảm bảo mức độ tin cậy chính xác.

e) Nguyên tắc tương tự của đối tượng dự báo

Nguyên tắc này đòi hỏi khi dự báo một đối tượng cần so sánh những tính chất của nó với những đối tượng tương tự đã biết và những mô hình có sẵn phục vụ cho việc dự báo đối tượng đó. Nguyên tắc này tiết kiệm được chi phí.

§2. PHƯƠNG PHÁP NGOẠI SUY

2.1. Giới thiệu về phương pháp

2.1.1. Khái niệm

Ngoại suy dự báo có nghĩa là nghiên cứu lịch sử của đối tượng và chuyển tính quy luật của nó đã phát hiện được trong quá khứ và hiện tại sang tương lai bằng các phương pháp xử lý dữ liệu chuỗi thời gian.

2.1.2. Dữ liệu chuỗi thời gian

Nếu quá trình ngẫu nhiên là một chuỗi các đại lượng ngẫu nhiên thì khi ta quan sát kết quả của *n* phép thử theo một đặc trưng nào đó thì chuỗi thời gian chính là một quá trình ngẫu nhiên.

Điều kiện dữ liệu chuỗi thời gian:

- Khoảng cách giữa các thời điểm của chuỗi thời gian phải bằng nhau, tức là phải đảm bảo tính liên tục nhằm phục vụ cho việc xử lý.

- Đơn vị đo giá trị chuỗi thời gian phải bằng nhau.

2.1.3. Cơ sở toán học và điều kiện

a) Cơ sở toán học

Ngoại suy cho ta thông tin tương lai của đối tượng dựa trên cơ sở chuỗi thời gian nhằm tìm ra xu thế của nó. Việc sử dụng ngoại suy trong dự báo dựa trên giả định: quá trình thay đổi của đối tượng thời gian là sự kết hợp các thành phần xu thế và thành phần ngẫu nhiên. Điều này dựa trên cơ sở toán học sau:

- Bản thân của dữ liệu chuỗi thời gian là một quá trình ngẫu nhiên.

- Quá trình ngẫu nhiên được mô tả bằng hàm ngẫu nhiên với sự tham gia của hai đại lượng.

+) Kỳ vọng: với nghĩa là hàm trung bình của hàm ngẫu nhiên mà những thực hiện ngẫu nhiên xoay quanh nó - $f(t)$

+) Phương sai: với đại diện cho sự phân tán của hàm ngẫu nhiên so với hàm trung bình - $\varepsilon(t)$

Khi đó $f(t)$ là thành phần xu thế cần xác định của đối tượng dự báo, nó nói lên ảnh hưởng của các nhân tố tác động thường xuyên trong một thời gian dài, còn $\varepsilon(t)$ là thành phần ngẫu nhiên biểu hiện những sai lệch so với thành phần xu thế, đây chính là tác động ngẫu nhiên của các nhân tố ngẫu nhiên đối với đối tượng dự báo.

Trên cơ sở đó ta có thể hoàn toàn dự báo theo công thức:

$$\hat{y}(t) = f(t) + \varepsilon(t)$$

b) Điều kiện

- Đối tượng phải phát triển tương đối ổn định theo thời gian.

- Những nhân tố ảnh hưởng chung nhất cho sự phát triển của đối tượng được duy trì trong khoảng thời gian dự báo.

- Không xảy ra đột biến trong quá trình phát triển của đối tượng.

2.2. Nội dung phương pháp

Phương pháp dự báo ngoại suy dựa trên dữ liệu chuỗi thời gian là một quá trình gồm nhiều giai đoạn quan trọng:

+ Xử lý chuỗi thời gian.

+ Phát hiện xu thế đối tượng.

+ Xây dựng hàm xu thế.

+ Xây dựng hàm xu thế và dự báo bằng hàm xu thế đã kiểm định.

2.2.1. Xử lý dữ liệu chuỗi thời gian

Chuỗi thời gian với các điều kiện của nó cần thiết được xử lý sơ bộ cho hoàn chỉnh.

a) Nếu chuỗi thiếu một giá trị y_t nào đó.

Lúc này ta xác định giá trị y_t bổ sung bằng trung bình cộng hai giá trị đứng trước và đứng sau nó:

$$y_t^{bs} = \frac{y_{t-1} + y_{t+1}}{2}$$

b) Xử lý dao động ngẫu nhiên

Việc phát hiện được ngay hàm xu thế $f(t)$ khi căn cứ vào chuỗi thời gian ban đầu là không dễ. Đối với chuỗi có dao động lớn do ảnh hưởng của các yếu tố ngẫu nhiên nên phải sử dụng các phương pháp làm trơn chuỗi thời gian với mục đích tạo ra một chuỗi thời gian mới có xu hướng dao động ổn định hơn, dĩ nhiên chuỗi thời gian tìm được chắc chắn vẫn giữ nguyên xu thế từ chuỗi thời gian.



Việc chuyển từ y_t sang $\overline{y}_t$ được xác định qua hai phương pháp san cơ bản sau:

* Trung bình trượt không có trọng số:

Phạm vi: áp dụng cho các chuỗi có khả năng tuân theo xu hướng đường thẳng (hàm tuyến tính bậc 1).

- Cách xác định:

$$\overline{y}_t = \frac{\sum_{i=t-p}^{t+p} y_i}{2p+1} \qquad (6.2)$$

trong đó: $m = 2p + 1$ là khoảng trượt;

y_i - giá trị chuỗi thời gian ban đầu vào thời điểm thứ I;

$\overline{y}_t$ - giá trị chuỗi thời gian được san vào thời điểm t.

Ví dụ 6.1. Có chuỗi thời gian về mức tiêu thụ sản phẩm x qua các năm như trong bảng 6.1.

Bảng 6.1. Bảng san chuỗi thời gian bằng phương pháp trung bình trượt không trọng số

t (năm)	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003
y_i (tấn)	3,4	4,1	3,7	3,8	4,5	4,7	5,1	3,9	4,2	4,1	4,8
$\overline{y}_t$ (m=3)	-	3,7	3,9	4,0	4,3	4,8	4,6	4,4	4,1	4,4	-

* Trung bình trượt có trọng số:

- Phạm vi: áp dụng cho các chuỗi số có xu hướng là đường cong (xu thế phi tuyến).

- Cách xác định: $\overline{y}_t$ của một khoảng trượt được mô tả bằng các đa thức bậc p.

$$\overline{y}_t = a_0 + \sum_{i=1}^{p} a_i t^i \qquad (6.3)$$

Các ký hiệu ($m = 2p + 1$), y_i, $\overline{y}_t$ vẫn được hiểu như trong phương pháp bình quân trượt không có trọng số.

Vấn đề là làm sao từ đa thức (6.3) ở trên có thể tìm ra mối quan hệ giữa $\overline{y}_t$ và y_i.

Đối với trường hợp đầu tiên ta xác định khoảng cách trượt $m = 5$ thì lúc này:

$$\overline{y}_t = a_0 + a_1 t + a_2 t^2 \;;\quad t \in \{-2,-1,0,1,2\} = \{t_1, t_2, t_3, t_4, t_5\} \qquad (6.4)$$

Theo phương pháp bình phương cực tiểu ta xác định được a_0, a_1, a_2 sao cho:

$$S(a_0,a_1,a_2)=\sum_{t=1}^{5}(y_t-\overline{y}_{t_t})^2 \rightarrow \min \leftrightarrow \sum_{t=1}^{5}(y_t-a_0-a_1t_t-a_2t^2)^2 \rightarrow \min$$

trong đó: y_t là giá trị thực tế của chuỗi thời gian.

$\overline{y}_{t_t}$ là giá trị lý thuyết của hàm xu thế $\overline{y}_t$.

Lấy đạo hàm của S theo các biến a_0, a_1, a_2 và cho bằng 0

$$\begin{cases} \dfrac{\partial S}{\partial a_0}=2\sum_{t=1}^{5}(y_t-a_0-a_1t_t-a_2t_t^2)^2(-1)=0 \\ \dfrac{\partial S}{\partial a_1}=2\sum_{t=1}^{5}(y_t-a_0-a_1t_t-a_2t_t^2)^2(-t_t)=0 \\ \dfrac{\partial S}{\partial a_2}=2\sum_{t=1}^{5}(y_t-a_0-a_1t_t-a_2t_t^2)^2(-t_t^2)=0 \end{cases} \quad (6.5)$$

$$\Leftrightarrow \begin{cases} \sum_{t=1}^{5}y_t=5a_0+a_1\sum_{t=1}^{5}t_t+a_2\sum_{t=1}^{5}t_t^2 \\ \sum_{t=1}^{5}y_tt_t=a_0\sum_{t=1}^{5}t_t+a_1\sum_{t=1}^{5}t_t^2+a_2\sum_{t=1}^{5}t_t^3 \\ \sum_{t=1}^{5}y_tt_t^2=a_0\sum_{t=1}^{5}t_t^2+a_1\sum_{t=1}^{5}t_t^3+a_2\sum_{t=1}^{5}t_t^4 \end{cases} \quad (6.6)$$

Lưu ý rằng : $\sum_{t=1}^{5}t_t=0$; $\sum_{t=1}^{5}t_t^2=10$; $\sum_{t=1}^{5}t_t^4=34$

Ta được hệ phương trình :

$$\begin{cases} \sum_{t=1}^{5}y_t=5a_0+10a_2 & (1) \\ \sum_{t=1}^{5}y_tt_t=10a_1 & (2) \\ \sum_{t=1}^{5}y_tt_t^2=10a_0+34a_2 & (3) \end{cases}$$

Từ (1), và (3) ta có : $a_o=\dfrac{1}{35}\left[17\sum_{t=1}^{5}y_t-5\sum_{t=1}^{5}y_tt_t^2\right]$ (4)

Mặt khác dễ thấy $\overline{y}_0=a_0$

Đánh số lại : $t_1 = t_{i-2}; t_2 = t_{i-1}; t_3 = t_i; t_4 = t_{i+1}; t_5 = t_{i+2}$

Thay giá trị của $t \in \{t_{i-2}, t_{i-1}, t_i, t_{i+1}, t_{i+2}\}$ vào (4) ta được :

$$\overline{y}_{t_i} = a_0 = \frac{1}{35}(-3y_{i-2} + 12y_{i-1} + 17y_i + 12y_{i+1} - 3y_{i+2})$$

Đây là công thức tổng quát để tính giá trị $\overline{y}_{t_i}$

2.2.2. Phát hiện hàm xu thế

Đây là giai đoạn quan trọng, mang tính quyết định đối với kết quả dự báo bằng phương pháp ngoại suy. Qua việc phân tích những nét chung nhất của chuỗi số liệu ta có thể phát hiện ra xu thế và hàm xu thế của đối tượng. Có nhiều phương pháp phát hiện xu thế và chọn hàm xu thế tương ứng, sau đây ta xét các phương pháp tiêu biểu được áp dụng rộng rãi.

a) Phương pháp đồ thị

Nội dung: biểu hiện các cặp số (t, y_i) lên hệ trục toạ độ, sau đó nối liền các điểm trên hệ trục thành một đường gẫy khúc liên tục, từ đó so sánh đường biểu diễn các hàm số $y = f(t)$ thường gặp hàm xác định xu thế và dạng hàm xu thế tương ứng.

Việc lựa chọn hàm xu thế theo phương pháp đồ thị phụ thuộc vào kinh nghiệm của người nghiên cứu, do đó rủi ro gặp phải là rất lớn.

b) Phương pháp phân tích số liệu quan sát

Nội dung: so sánh dữ liệu của đối tượng với một số dạng hàm cơ bản , từ đó tìm ra dạng hàm xu thế thích hợp.

1. Dạng $\hat{y}_t = a_0 + a_1 t$

$$t_i \approx \frac{t_{i-1} + t_{i+1}}{2} \quad \text{và} \quad y_i \approx \frac{y_{i-1} + y_{i+1}}{2}$$

2. Dạng $\hat{y}_t = a_0 a_1^t$ điều kiện

$$t_i \approx \frac{t_{i-1} + t_{i+1}}{2} \quad \text{và} \quad y_i = \sqrt{y_{i-1} \cdot y_{i+1}}$$

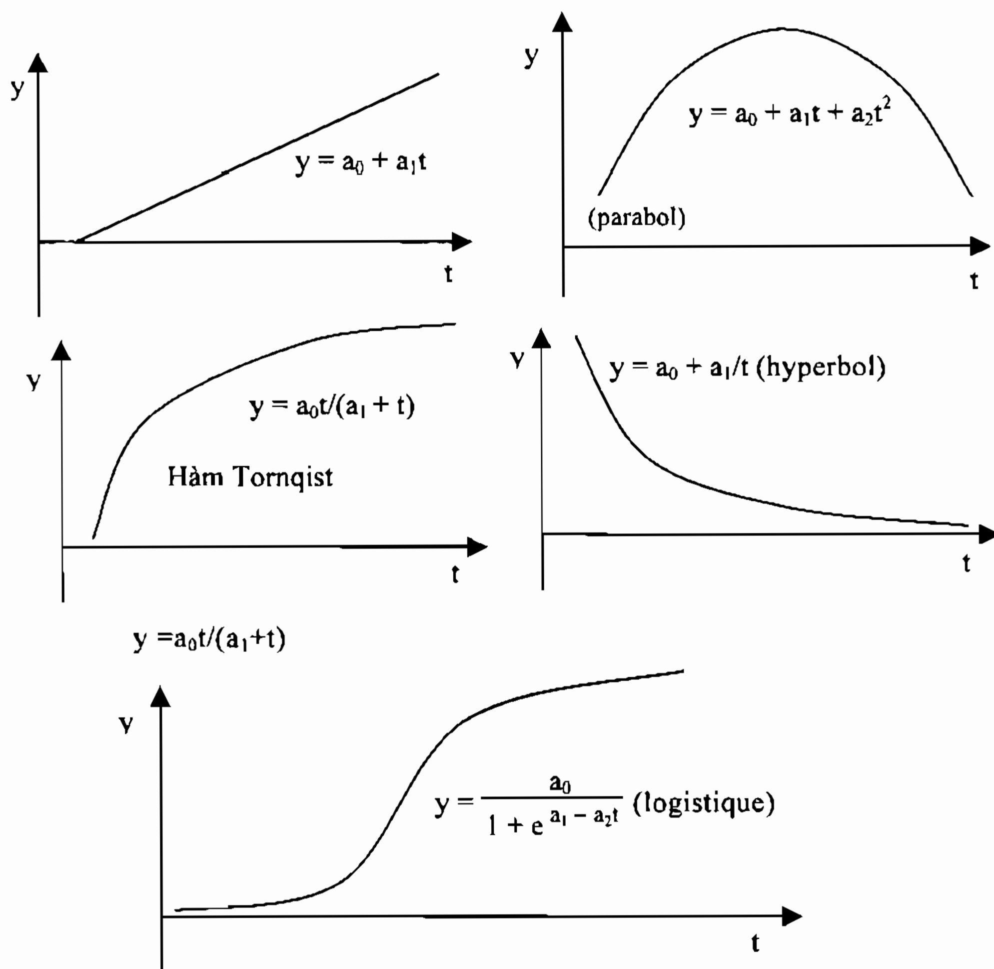
3. Dạng $\hat{y}_t = a_0 t^{a_1}$ điều kiện

$$\ln t_i \approx \frac{\ln t_{i-1} + \ln t_{i+1}}{2} \quad \text{và} \quad \ln y_i = \frac{\ln y_{i-1} + \ln y_{i+1}}{2}$$

c) Phương pháp sai phân

Nội dung: phương pháp dựa trên cơ sở sự xấp xỉ giữa sai phân chuỗi thời

gian và vi phân hàm xu thế ở cùng bậc k nào đó ($\Delta_y^k \approx d_y^k$). Do vậy, chúng ta có thể lấy sai phân bậc k của chuỗi thời gian, nếu dừng lại ở bậc sai phân nào đó mà các giá trị sai phân đều có xu hướng tiến về hằng số thì kết luận có khả năng thích hợp với dạng hàm xu thế.



Hình 6.2. Phát hiện hàm xu thế bằng phương pháp đồ thị.

$$\hat{y} = a_0 + \sum_{i=1}^{k} a_1 t^i$$

Sai phân chuỗi thời gian được định nghĩa như sau:

- Sai phân bậc nhất : $\Delta y_{t+1} = y_{t+1} - y_t$

- Sai phân bậc hai: $\Delta y_{i+1}^2 = \Delta y_{i+1} - \Delta y_i$

- Sai phân bậc ba: $\Delta y_{i+1}^3 = \Delta y_{i+1}^2 - \Delta y_i^2$

- Sai phân bậc k: $\Delta y_{i+1}^k = \Delta y_{i+1}^{k-1} - \Delta y_i^{k-1}$

Vi phân các hàm số dạng: $\hat{y} = a_0 + \sum_{i=1}^{k} a_i t^i$ cho trong bảng 6.2

Bảng 6.2. Dạng vi phân của hàm bậc hai và bậc ba

Hàm $\hat{y}$	$d\hat{y}/dt$	$d^2\hat{y}/dt^2$	$d^3\hat{y}/dt^3$
$\hat{y} = a_0 + a_1 t$	a_1	0	0
$\hat{y} = a_0 + a_1 t + a_2 t^2$	$a_1 + 2a_2 t$	$2a_2$	0
$\hat{y} = a_0 + a_1 t + a_2 t^2 + a_3 t^3$	$a_1 + 2a_2 t + 3a_3 t^2$	$2a_2 + 6a_3 t$	$6a_3$

Phương pháp này chỉ áp dụng cho các hàm đa thức có dạng tổng quát là: $y = a_0 + \sum_{i=1}^{k} a_i t^i$. Các hàm phi tuyến khác sẽ không có giá trị vì khi càng nâng cao bậc thì càng trở nên phức tạp và không có cơ sở so sánh với sai phân tương ứng.

2.2.3. Xây dựng hàm xu thế

Khi phát hiện ra các khả năng về dạng hàm xu thế, vấn đề kế tiếp phải mô tả chuỗi thời gian thông qua các dạng hàm xu thế cụ thể với điều kiện xác định những tham số a_i; i = 1, 2, ..., k của nó với những giá trị bằng số cụ thể. Thường áp dụng các phương pháp sau:

a) Phương pháp điểm chọn

Đây là một phương pháp đơn giản xác định được các tham số a_i ở mức độ xấp xỉ, tuy nhiên nó có nhược điểm là lãng phí thông tin và tuỳ thuộc vào cách chọn các điểm chọn mà ta có các bộ tham số a_i khác nhau.

Nội dung: phương pháp này giả định hàm lý thuyết đã chọn ở bước phát hiện xu thế, chọn các cặp số (t_i, y_i).

Yêu cầu:

- Khoảng cách giữa các điểm được chọn bằng nhau,

- Tổng số các điểm chọn bằng tổng số các tham số a_i.

- Do yêu cầu về độ chính xác cần chọn những điểm thực nghiệm, mà đường biểu diễn của hàm xu thế có khả năng đi qua cao nhất

Ví dụ 6.2: có chuỗi thời gian (bảng 6.3)

Bảng 6.3: Bảng chuỗi thời gian cho ví dụ 6.2

t_i	1	2	3	4	5	6	7	8	9	10	11	12
y_i	1	3	6	5	8	7	6	9	16	25	19	32

Ở giai đoạn phát hiện xu thế ta kết luận hàm xu thế có dạng: $\hat{y} = a_0 + a_1 t + a_2 t^2$.

Bằng phương pháp điểm chọn, xác định các tham số a_i.

Chọn 3 điểm $t = 2$; $t = 6$; $t = 10$. Nhìn vào bảng (6.3) ta tìm được các điểm $y_2 = 2$; $y_6 = 7$; $y_{10} = 25$. Ta có:

$$\begin{cases} a_0 + 2a_1 + 4a_2 = 2 \\ a_0 + 6a_1 + 36a_2 = 7 \\ a_0 + 10a_1 + 100a_2 = 25 \end{cases} \Leftrightarrow \begin{cases} a_0 = 4,375 \\ a_1 = -2 \\ a_2 \approx 0,406 \end{cases}$$

Vậy hàm xu thế mô tả cho chuỗi thời gian có dạng: $\hat{y} = 4,375 - 2t - 0,406t^2$.

b) Phương pháp tổng bình phương bé nhất.

Là một trong những phương pháp được ứng dụng rộng rãi nhất để xác định các tham số của hàm xu thế, mức độ chính xác của nó thể hiện ở chỗ: tổng bình phương độ lệch giữa giá trị lý thuyết hàm xu thế và giá trị thực tế của chuỗi thời gian là nhỏ nhất.

$$S = \sum_{i=1}^{n} (y_i - \hat{y})^2 \rightarrow \min$$

trong đó n: - độ dài của chuỗi thời gian .

y_i - giá trị thực tế của chuỗi thời gian; $i = 1, 2, .., n$.

$\hat{y}$ - giá trị lý thuyết của hàm xu thế.

Theo lý thuyết, để S đạt min ta lấy đạo hàm bậc nhất của S theo các tham số a_i; $i = 1, 2, \ldots, n$. Sau đó cho các biểu thức bằng 0 và giải hệ phương trình vừa tạo ra, ta thu được các nghiệm là các tham số a_i. Cụ thể:

+) Nếu hàm xu thế có dạng hàm bậc nhất: $\hat{y} = a_0 + a_1 t$

Ta có : $S=\sum_{i=1}^{n}(y_i-a_0-a_1t)^2$ (6.7)

Lấy đạo hàm bậc nhất (6.7) và cho các biểu thức bằng 0 ta được :

$$\begin{cases}\dfrac{\partial S}{\partial a_0}=\sum_{i=1}^{n}2(y_i-a_0-a_1t)(-1)=0\\ \dfrac{\partial S}{\partial a_1}=\sum_{i=1}^{n}2(y_i-a_0-a_1t)(-t)=0\end{cases}\Leftrightarrow\begin{cases}\sum_{i=1}^{n}y_i=na_0+a_1\sum_{i=1}^{n}t_i\\ \sum_{i=1}^{n}y_it_i=a_0\sum_{i=1}^{n}t_i+a_1\sum_{i=1}^{n}t_i^2\end{cases}$$

Giải hệ trên ta được 2 nghiệm là giá trị của a_0 và a_1.

Nếu hàm xu thế có dạng bậc 2: $\hat{y}=a_0+a_1t+a_2t^2$

Ta có $S=\sum_{i=1}^{n}(y_i-a_0-a_1t-a_2t^2)^2\rightarrow$ min (6.8)

$\Leftrightarrow$ Lấy đạo hàm bậc nhất (6.8) và cho các biểu thức bằng 0 ta được :

$$\begin{cases}\dfrac{\partial S}{\partial a_0}=\sum_{i=1}^{n}2(y_i-a_0-a_1t-a_2t^2)(-1)=0\\ \dfrac{\partial S}{\partial a_1}=\sum_{i=1}^{n}2(y_i-a_0-a_1t-a_2t^2)(-t)=0\\ \dfrac{\partial S}{\partial a_2}=\sum_{i=1}^{n}2(y_i-a_0-a_1t-a_2t^2)(-t^2)=0\end{cases}$$

$$\Leftrightarrow\begin{cases}\sum_{i=1}^{n}y_i=na_0+a_1\sum_{i=1}^{n}t_i+a_2\sum_{i=1}^{n}t_i^2\\ \sum_{i=1}^{n}y_it_i=a_0\sum_{i=1}^{n}t_i+a_1\sum_{i=1}^{n}t_i^2+a_2\sum_{i=1}^{n}t_i^3\\ \sum_{i=1}^{n}y_it_i^2=a_0\sum_{i=1}^{n}t_i^2+a_1\sum_{i=1}^{n}t_i^3+a_2\sum_{i=1}^{n}t_i^4\end{cases}\qquad(6.9)$$

Giải hệ phương trình trên ta được các nghiệm là các giá trị là các tham số a_0, a_1, a_2.

Tổng quát: hàm xu thế có dạng: $\hat{y}=a_0+\sum_{j=1}^{p}a_jt^j$

Khi đó: $S=\sum_{i=1}^{n}(y_i-a_0-\sum_{j=1}^{p}a_jt^j)^2\rightarrow$ min (6.10)

$\Leftrightarrow$ Lấy đạo hàm bậc nhất và cho các biểu thức bằng 0. Ta có

$$\begin{cases} \dfrac{\partial S}{\partial a_0} = 0 \\ \dfrac{\partial S}{\partial a_1} = 0 \\ \cdots\cdots \\ \dfrac{\partial S}{\partial a_p} = 0 \end{cases} \Leftrightarrow \begin{cases} \sum_{i=1}^{n} y_i = na_0 + a_1\sum_{i=1}^{n} t_i + a_2\sum_{i=1}^{n} t_i^2 + \ldots + a_p\sum_{i=1}^{n} t_i^p \\ \sum_{i=1}^{n} y_i t_i = a_0\sum_{i=1}^{n} t_i + a_1\sum_{i=1}^{n} t_i^2 + a_2\sum_{i=1}^{n} t_i^3 + \ldots + a_p\sum_{i=1}^{n} t_i^{p+1} \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \sum_{i=1}^{n} y_i t_i^p = a_0\sum_{i=1}^{n} t_i^p + a_1\sum_{i=1}^{n} t_i^{p+1} + a_2\sum_{i=1}^{n} t_i^{p+2} + \ldots + a_p\sum_{i=1}^{n} t_i^{2p} (j = \overline{1,p}) \end{cases}$$

Giải hệ phương trình trên ta thu được các giá trị của các tham số a_i.

Nhận xét: phương pháp tổng bình phương bé nhất tạo ra các tham số a_i với độ chính xác cao hơn nhiều so với các phương pháp khác.

Phương pháp này chỉ áp dụng cho các dạng hàm $\hat{y} = a_0 + \sum_{i=1}^{p} a_i t^i$, nếu muốn áp dụng cho các dạng hàm phi tuyến thì trước tiên phải tuyến tính hoá chúng.

Trong quá trình giải hệ phương trình tìm tham số a_i của hàm xu thế, để đơn giản và rút gọn qúa trình tính toán ta có thể thiết kế một quy luật t_i sao cho: $\sum^{n}_{t_i chon} = 0$.

+) Nếu n là số chẵn, ta chọn $t_{chọn}$ như sau:

t_i (ban đầu)	1	2	3	4	5	6	7	8	9	10
t_i (chọn)	-5	-3	-1	1	3	5	7	9	11	13
	Quá khứ				Hiện tại		Tương lai			

+) Nếu n là lẻ, ta chọn $t_{chọn}$ như sau:

t_i (ban đầu)	1	2	3	4	5	6	7	8	9	10
t_i (chọn)	-3	-2	-1	0	1	2	3	4	5	6
	Quá khứ					Hiện tại		Tương lai		

Nhược điểm của phương pháp:

Trong trường hợp các tham số của hàm xu thế phải thay đổi theo thời gian cho phù hợp với xu thế của đối tượng, thì việc áp dụng phương phấp tổng bình phương bé nhất có thể dẫn đến sai lầm cho kết quả dự báo.

Ví dụ 6.3:

Giả sử số lượng sản lượng bán được của một doanh nghiệp trong 10 tuần đầu có hàm xu thế là y = 1500 + 10*t*. Sau khi đưa ra chiến dịch tiếp thị hợp lý số lượng sản phẩm bán ra ngày càng tăng lên. Khi đó nếu vẫn áp dụng hàm xu thế ban đầu thì sẽ dự báo sai số lượng sản phẩm bán được trong các tuần tiếp theo. Thay vào đấy ta có thể sử dụng hàm xu thế sau: y = 1550 + 20t để thu được kết quả dự báo chính xác hơn.

2.2.4. Kiểm định hàm xu thế và dự báo

a) Kiểm định hàm xu thế

Do ở bước phát triển hàm xu thế, hàm phát được chỉ mang tính chất khả năng, vì vậy phải tiến hành kiểm định nhằm đánh giá lựa chọn hàm xu thế tối ưu.

*) Các tiêu chí kiểm định:

- Sai số tuyệt đối: $S_y = \sqrt{\dfrac{\sum_{t=1}^{n}(\hat{y}_t - y_t)^2}{n-2}}$

- Sai số tương đối: $V_{y\%} = \dfrac{S_y}{\bar{y}}.100 = \dfrac{S_y}{\frac{1}{n}\sum_{t=1}^{n} y_t}$

y - giá trị thực tế.

$\hat{y}$ - giá trị lý thuyết hàm xu thế.

n - số mức độ của chuỗi.

*) Giới hạn lựa chọn hàm xu thế.

- Nếu V_y > 10% thì hàm xu thế không được sử dụng cho dự báo.

- Nếu $V_y \leq$ 10% thì hàm xu thế được sử dụng cho dự báo.

b) Dự báo bằng hàm xu thế đã kiểm định

*) Dự báo điểm:

Khoảng cách dự báo l: để đảm bảo tính chính xác ta chỉ lấy khoảng l sao cho:

$$l_{max} \leq \frac{n}{3}$$

Dự báo với khoảng cách dự báo l được xác định như sau :

$$\hat{y}_{n+1}^{DB} = f(n+l)$$

*) Dự báo khoảng:

Là dự báo rơi vào khoảng nhất định với xác suất cho trước

$$\hat{y}^{DH}_{n+l} = [f(n+l) - t_\alpha S_l ; f(n+l) + t_\alpha S_l]$$

trong đó: t_α là giá trị của hàm phân phối Student với $(n-p)$ bậc tự do, và độ tin cậy là $(1-\alpha)$.

S_l là sai số dự báo, S_l được tính như sau:

1. Nếu xu thế là hàm tuyến tính bậc nhất :

$$S_l = S_y\sqrt{1 + \frac{1}{n} + \frac{3(n+2l-1)^2}{n(n^2-1)}} \quad \text{với sai số tuyệt đối } S_y = \sqrt{\frac{\sum_{t=1}^{n}(\hat{y}-y)^2}{n-2}}$$

2. Nếu xu thế là hàm bậc 2, bậc 3:

$$S_l = S_y\sqrt{1 + T_l(T^{-1})T_l^T} \quad \text{với } S_y \text{ được tính như trên.}$$

T_l là vector dòng các giá trị luỹ thừa của t tại thời điểm $n+l$.

T^{-1} là ma trận nghịch đảo của ma trận hệ phương trình chuẩn.

Xác định T_l và T^{-1}

Hàm bậc 2	Hàm bậc 3
$T_l = (1, t_l, t_l^2)$ $T^{-1} = \begin{bmatrix} n & \sum_{t=1}^{n} t & \sum_{t=1}^{n} t^2 \\ \sum_{t=1}^{n} t & \sum_{t=1}^{n} t^2 & \sum_{t=1}^{n} t^3 \\ \sum_{t=1}^{n} t^2 & \sum_{t=1}^{n} t^3 & \sum_{t=1}^{n} t^4 \end{bmatrix}$	$T_l = (1, t_l, t_l^2, t_l^3)$ $T^{-1} = \begin{bmatrix} n & \sum_{t=1}^{n} t & \sum_{t=1}^{n} t^2 & \sum_{t=1}^{n} t^3 \\ \sum_{t=1}^{n} t & \sum_{t=1}^{n} t^2 & \sum_{t=1}^{n} t^3 & \sum_{t=1}^{n} t^4 \\ \sum_{t=1}^{n} t^2 & \sum_{t=1}^{n} t^3 & \sum_{t=1}^{n} t^4 & \sum_{t=1}^{n} t^5 \\ \sum_{t=1}^{n} t^3 & \sum_{t=1}^{n} t^4 & \sum_{t=1}^{n} t^5 & \sum_{t=1}^{n} t^6 \end{bmatrix}$

3. Các dạng còn lại: $S_l = S_y$

Ví dụ 6.4. Có số liệu về sự biến động của sản xuất (tính theo sản lượng) về một loại sản phẩm x các tháng trong năm 2005 ở bảng 6.4 sau:

Bảng 6.4

Tháng	1	2	3	4	5	6	7	8	9	10	11	12
(1000 chiếc)	100	150	195	220	260	280	300	310	300	270	260	240

Sau khi qua bước phát triển hàm xu thế, chuỗi số liệu trên thích hợp với hàm dạng: $y = a_0 + a_1 t + a_2 t^2$. Nếu số mức độ của chuỗi thời gian là số lẻ. Ta sẽ xác định $t_{chọn}$ như bảng 6.5.

Bảng 6.5

t_i (đầu)	1	2	3	4	5	6	7	8	9	10	11	12
y_i	100	150	195	220	260	280	300	310	300	270	260	240
t_i (sau)	-4	-3	-2	-1	0	1	2	3	4	5	7	9
y_i												

Ta sẽ tìm a_0, a_1, a_2 để $S(a_0, a_1, a_2) = \sum_{i=-4}^{4}(y_i - \hat{y}_i) \to \min$,

trong đó $\hat{y}_i = a_0 + a_1 t_i + a_2 t_i^2$; $i = \overline{-4,4}$.

$$\frac{\partial S}{\partial a_j} = 2\sum_{i=-4}^{4}(y_i - a_0 - a_1 t_i - a_2 t_i^2)(-t_i^j); j = 0,1,2$$

Ta có hệ phương trình sau đây:

$$\begin{cases} \dfrac{\partial S}{\partial a_0} = 0 \\ \dfrac{\partial S}{\partial a_1} = 0 \\ \dfrac{\partial S}{\partial a_p} = 0 \end{cases} \Leftrightarrow \begin{cases} \sum_{i=-4}^{4}(y_i - a_0 - a_1 t_i - a_2 t_i^2) = 0 & (1) \\ \sum_{i=-4}^{4}(y_i - a_0 - a_1 t_i - a_2 t_i^2)(-t_i) = 0 & (2) \\ \sum_{i=-4}^{4}(y_i - a_0 - a_1 t_i - a_2 t_i^2)(-t_i^2) = 0 & (3) \end{cases}$$

Thay số vào ta thấy (1) $\Leftrightarrow$ $9a_0 + 6a_2 = 2115$ (1')

(2) $\Leftrightarrow$ $60a_1 = 1550$ (2')

(3) $\Leftrightarrow$ $60a_0 + 708a_2 = 13020$ (3')

Giải hệ gồm các phương trình (1'),(2'),(3'). Ta được

$a_0 = 258,4$; $a_1 = 25,8$; $a_0 = -3,51$

Vậy $\hat{y} = 258,4 + 25,8t - 3,51t^2$

Kiểm định:

$$\bar{y} = \frac{1}{9}\sum_{t=-4}^{4} y_t = 235; \quad S_y = \sqrt{\frac{\sum_{t=-4}^{4}(\hat{y}_t - y_t)^2}{n-2}} = \sqrt{\frac{167,25}{7}} = 4,888$$

$$V_{y\%} = \frac{S_y}{\bar{y}}.100 = \frac{4,888}{235}100\% = 2\% < 10\%.$$

Vậy mô hình dự báo đưa ra là chấp nhận được.

VÀI DÒNG KẾT LUẬN

Khi nghiên cứu Kinh tế lượng ta sẽ thực sự thấy say mê và hứng thú. Việc tìm hiểu phương pháp luận của Kinh tế lượng giúp chúng ta tìm được mô hình tổng quát của một hoạt động kinh tế dựa trên những bộ số liệu đầu vào, tìm ra mối quan hệ giữa các số liệu đầu vào, phân tích đánh giá để tìm ra mô hình đúng đắn và sát thực cho hoạt động kinh tế. Thông qua đó, ta sẽ biết quá trình vận động, quá trình tương tác lẫn nhau giữa các hiện tượng kinh tế, cũng như các hiện tượng khoa học tự nhiên và khoa học xã hội.

Sự đòi hỏi của Kinh tế lượng trong phân tích định lượng các các hiện tượng kinh tế, kiểm định sự phù hợp và độ tin cậy của các giả thiết trong quá trình hoạch định các chính sách vĩ mô cũng như ra các quyết định tác nghiệp, việc dự báo và dự đoán có độ tin cậy cao,... , tất cả những điều đó đã làm cho Kinh tế lượng có một vai trò ngày càng quan trọng và bản thân nó cũng không ngừng được phát triển và hoàn thiện.

Tóm lại, qua việc nghiên cứu Kinh tế lượng, ta sẽ thấy được:

1. Kinh tế lượng đã đem lại một công cụ đo lường sắc bén để đo lường các quan hệ kinh tế.

2. Trang bị cho nhà kinh tế một phương pháp và phân tích sự vận động của các hoạt động kinh tế.

3. Qua việc nghiên cứu Kinh tế lượng, chúng ta đưa ra được một mô hình dưới dạng phù hợp, ước lượng sau đó kiểm định bằng thực nghiệm một cách chính xác đưa ra dự báo với độ tin cậy cao dẫn tới những quyết định đúng đắn trong hoạt động kinh doanh tác nghiệp, hoạch định các chính sách và chiến lược kinh tế xã hội.

PHỤ LỤC

PHỤ LỤC 1

CÁC BẢNG THỐNG KÊ

Bảng 1. BẢNG STUDENT

v	t_{55}	t_{60}	t_{70}	t_{75}	t_{80}	t_{90}	t_{95}	t_{975}	t_{99}	t_{995}
1	,158	,325	,727	1,000	1,376	3,08	9,31	12,71	31,82	63,66
2	,142	,298	,617	,816	1,061	1,89	2,92	4,30	6,96	9,92
3	,137	,277	,584	,765	,978	1,64	2,35	3,18	4,54	5,84
4	,134	,271	,569	,741	,941	,153	2,13	2,78	3,75	4,60
5	,132	,267	,559	,727	,920	1,48	2,02	2,57	3,36	4,03
6	,131	,265	,553	,718	,906	1,44	1,94	2,45	3,14	3,71
7	,130	,263	,549	,711	,896	1,42	1,90	2,36	3,00	3,50
8	,130	,262	,546	,706	,889	1,40	1,86	2,31	2,90	3,36
9	,129	,261	,543	,703	,883	1,38	1,83	2,26	2,82	3,25
10	,129	,260	,542	,700	,879	1,37	1,81	2,23	2,76	3,17
11	,129	,260	,540	,697	,876	1,36	1,80	2,20	2,72	3,11
12	,128	,259	,539	,695	,873	1,36	1,78	2,18	2,68	3,06
13	,128	,259	,538	,694	,870	1,35	1,77	2,16	2,65	3,01
14	,128	,258	,537	,692	,868	1,34	1,76	2,14	2,62	2,98
15	,128	,258	,536	,691	,866	1,34	1,75	2,13	2,60	2,95
16	,128	,258	,535	,691	,865	1,34	1,75	2,12	2,58	2,92
17	,128	,257	,534	,689	,863	1,33	1,74	2,11	2,57	2,90
18	,127	,257	,534	,688	,862	1,33	1,73	2,19	2,55	2,88
19	,127	,257	,533	,688	,861	1,33	1,73	2,09	2,54	2,86
20	,127	,257	,533	,687	,860	1,32	1,72	2,09	2,53	2,84
21	,127	,257	,532	,868	,859	1,32	1,72	2,08	2,52	2,83
22	,127	,256	,532	,686	,858	1,32	1,72	2,07	2,51	2,82
23	,127	,256	,532	,865	,858	1,32	1,71	2,07	2,50	2,81
24	,127	,256	,531	,685	,857	1,32	1,71	2,06	2,49	2,80
25	,127	,256	,531	,684	,856	1,32	1,71	2,06	2,48	2,79
26	,127	,256	,531	,684	,856	1,32	1,71	2,06	2,48	2,78
27	,127	,256	,531	,684	,855	1,31	1,79	2,05	2,47	2,77
28	,127	,256	,530	,683	,855	1,31	1,70	2,05	2,47	2,76
29	,127	,256	,530	,683	,854	1,31	1,79	2,04	2,46	2,76

Bảng 1. (tiếp theo)

v	t_{55}	t_{60}	t_{70}	t_{75}	t_{80}	t_{90}	t_{95}	t_{975}	t_{99}	t_{995}
30	,127	,256	,530	,683	,854	1,31	1,70	2,04	2,46	2,75
40	,126	,255	,529	,681	,851	1,30	1,68	2,02	2,42	2,70
60	,126	,254	,527	,679	,848	1,30	1,67	2,00	2,39	2,66
120	,126	,254	,526	,677	,845	1,29	1,66	1,98	2,36	2,62
∞	,126	,253	,524	,674	,842	1,28	1,645	1,96	2,33	2,58

Source. R. A. Fisher and F. Yates, *Statiscal Tables for Biologica, Agricutural and Medical Research*, Longman Group Ltd, London (publides auparavant par Olivier and Boys, Edinburgh). Reproduit avec l'autorisation de Pauteur et des ditcurs.

Bảng 2. BẢNG THỐNG KÊ DURBIN - WATSON

X variables, excluding the intercept

Observation		1		2		3		4		5	
N	Prob.	D-L	D-U	D-L	D-U	D-L	D-U	D-L	D-U	D-L	D-U
15	0,05	1,08	1,36	0,95	1,54	0,82	1,75	0,69	1,97	0,56	2,21
	0,01	0,81	1,07	0,70	1,25	0,59	1,46	0,49	1,70	0,39	1,96
20	0,05	1,20	1,71	1,10	1,54	1,00	1,68	0,90	1,83	0,79	1,99
	0,01	0,95	1,15	0,86	1,27	0,77	1,41	0,68	1,57	0,60	1,74
25	0,05	1,29	1,45	1,21	1,55	1,12	1,66	1,04	1,77	0,95	1,89
	0,01	1,05	1,21	0,98	1,30	0,90	1,41	0,83	1,52	0,75	1,65
30	0,05	1,35	1,49	1,28	1,57	1,21	1,65	1,14	1,74	1,07	1,83
	0,01	1,13	1,26	1,07	1,34	1,01	1,42	0,94	1,51	0,88	1,61
40	0,05	1,44	1,54	1,39	1,60	1,34	1,66	1,39	1,72	1,23	1,79
	0,01	1,25	1,34	1,20	1,40	1,15	1,46	1,10	1,52	1,05	1,58
50	0,05	1,50	1,59	1,46	1,63	1,42	1,67	1,38	1,72	1,34	1,77
	0,01	1,32	1,40	1,28	1,45	1,24	1,49	1,20	1,54	1,16	1,59
60	0,05	1,55	1,62	1,51	1,65	1,48	1,69	1,44	1,73	1,41	1,77
	0,01	1,38	1,45	1,35	1,48	1,32	1,52	1,28	1,56	1,25	1,60
80	0,05	1,61	1,66	1,59	1,69	1,56	1,72	1,53	1,74	1,51	1,77
	0,01	1,47	1,52	1,44	1,54	1,42	1,57	1,39	1,60	1,36	1,62
100	0,05	1,65	1,69	1,63	1,72	1,61	1,74	1,59	1,76	1,57	1,78
	0,01	1,52	1,56	1,50	1,58	1,48	1,60	1,46	1,63	1,44	1,65

Bảng 3. BẢNG FISHER

$\alpha = 0,05$

n_1 \ n_2	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	161	200	216	225	230	234	237	239	241	242	244	246	248	249	250	251	252	253	254
2	18,5	19,0	19,2	19,2	19,3	19,4	19,4	19,4	19,4	19,4	19,4	19,4	19,4	19,5	19,5	19,5	19,5	19,5	19,5
3	10,1	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,74	8,70	8,66	8,64	8,62	8,59	8,57	8,55	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,91	5,86	5,80	5,77	5,75	5,72	5,69	5,66	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,68	4,62	4,56	4,53	4,50	4,46	4,43	4,40	4,37
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,00	3,94	3,87	3,84	3,81	3,77	3,74	,370	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,57	3,51	3,44	3,41	3,38	3,34	3,30	3,27	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,28	3,22	3,15	3,12	3,08	3,04	3,01	2,97	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,07	3,01	2,94	2,90	2,86	2,83	2,79	2,75	2,71
10	4,96	4,10	2,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,91	2,85	2,77	2,74	2,70	2,66	2,62	2,58	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,79	2,72	2,65	2,61	2,57	2,53	2,49	2,45	2,40
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,69	2,62	2,54	2,51	2,47	2,43	2,38	2,34	2,30
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,60	2,53	2,46	2,42	2,38	2,34	2,30	2,25	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,69	2,53	2,46	2,39	2,35	2,31	2,27	2,22	2,18	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,48	2,40	2,33	2,29	2,25	2,20	2,16	2,11	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,42	2,35	2,28	2,24	2,19	2,15	2,11	2,06	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,60	2,55	2,49	2,45	2,38	2,31	2,23	2,19	2,15	2,10	2,06	2,01	1,96

Bảng 3 (tiếp theo)

n_1 \ n_2	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,34	2,27	2,19	2,15	2,11	2,06	2,02	1,97	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,31	2,23	2,16	2,11	2,07	2,03	1,98	1,93	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,28	2,20	2,12	2,08	2,04	1,99	1,95	1,90	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,25	2,18	2,10	2,05	2,01	1,96	1,92	1,87	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,23	2,15	2,07	2,03	1,98	1,94	1,89	1,84	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27	2,20	2,13	2,05	2,01	1,96	1,91	1,86	1,81	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,18	2,11	2,03	1,98	1,94	1,89	1,84	1,79	1,73
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24	2,16	2,09	2,01	1,96	1,92	1,87	1,82	1,77	1,71
26	4,23	2,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,15	2,07	1,99	1,95	1,90	1,85	1,80	1,75	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25	2,20	2,13	2,06	1,97	1,93	1,88	1,84	1,79	1,73	1,67
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19	2,12	2,04	1,96	1,91	1,87	1,82	1,77	1,71	1,65
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22	2,18	2,10	2,03	1,94	1,90	1,85	1,81	1,75	1,70	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,09	2,01	1,93	1,89	1,84	1,79	1,74	1,68	1,62
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	2,00	1,92	1,84	1,79	1,74	1,69	1,64	1,58	1,51
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,92	1,84	1,75	1,70	1,65	1,59	1,53	1,47	1,39
120	3,92	3,07	2,68	2,45	2,29	2,18	2,09	2,02	1,96	1,91	1,83	1,75	1,66	1,61	1,55	1,50	1,43	1,35	1,25
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83	1,75	1,67	1,57	1,52	1,46	1,39	1,32	1,22	1,00

Source: E.S. Pearson and H.O. Hartley, *Biometrika Tables for Statisticians*, Vol.2 (1972), table 5, pages 178. Reprodution autoride

Bảng 4. BẢNG FISHER

$\alpha = 0,01$

n_2 \ n_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	4052	5000	5403	5625	5764	5859	5928	5981	6023	6056	6160	6157	6209	6235	6261	6287	6313	6339	6356
2	98,5	99,0	99,2	99,2	99,3	99,3	99,4	99,4	99,4	99,4	99,4	99,4	99,4	99,5	99,5	99,5	99,5	99,5	99,5
3	34,1	30,8	29,5	28,7	28,2	27,9	27,7	27,5	27,3	27,2	27,1	26,9	26,7	26,6	26,5	26,4	26,3	26,2	26,1
4	21,2	18,0	16,7	16,0	15,5	15,2	15,0	14,8	14,7	14,5	14,4	14,2	14,0	13,9	13,8	13,7	13,7	13,6	13,5
5	16,3	13,3	12,1	11,4	11,0	10,7	10,5	10,3	10,2	10,1	9,89	9,72	9,55	9,47	9,38	9,29	9,20	9,11	9,02
6	13,7	10,9	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,72	7,56	7,40	7,31	7,23	7,14	7,06	6,97	6,88
7	12,2	9,55	8,45	7,85	7,46	7,19	6,99	6,84	6,72	6,62	6,47	6,31	3,16	3,07	5,99	5,91	5,82	5,74	5,65
8	11,3	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,91	5,81	5,67	5,52	5,36	5,28	5,20	5,12	5,03	4,95	4,86
9	10,6	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,35	5,26	5,11	4,96	4,81	4,73	4,65	4,57	4,48	4,40	4,31
10	10,0	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,94	4,85	4,71	4,56	4,41	4,33	4,25	4,17	4,08	4,00	3,91
11	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,63	4,54	4,40	4,25	4,10	4,02	3,94	3,86	3,78	3,69	3,60
12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,39	4,30	4,16	4,01	3,86	3,78	3,70	3,62	3,54	3,45	3,36
13	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,30	4,19	4,10	3,96	3,82	3,66	3,59	3,51	3,43	3,34	3,25	3,17
14	8,86	6,51	5,56	5,04	4,70	4,46	4,28	4,14	4,03	3,94	3,80	3,66	3,51	3,43	3,35	3,27	3,18	3,09	3,00
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80	3,67	3,52	3,37	3,29	3,21	3,13	3,05	2,96	2,87
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69	3,55	3,41	3,26	3,18	3,10	3,02	2,93	2,84	2,75
17	8,40	6,11	5,19	4,67	4,34	4,10	3,93	3,79	3,68	3,59	3,46	3,31	3,16	3,08	3,00	2,92	2,83	2,75	2,65

Bảng 4 (tiếp theo)

n_2 \ n_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
18	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,71	3,60	3,51	3,37	3,23	3,08	3,00	2,92	2,84	2,75	2,66	2,57
19	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52	3,43	3,30	3,15	3,00	2,92	2,84	2,76	2,67	2,58	2,49
20	8,10	5,58	4,94	4,43	4,10	3,87	3,70	3,56	3,46	3,37	3,23	3,09	2,94	2,86	2,78	2,69	2,61	2,52	2,42
21	8,02	5,78	4,87	4,37	4,04	3,81	3,64	3,51	3,40	3,31	3,17	3,03	2,94	2,86	2,78	2,69	2,61	2,52	2,42
22	7,95	5,72	4,82	4,31	3,99	3,76	3,59	3,45	3,35	3,26	3,12	2,98	2,83	2,75	2,67	2,58	2,50	2,40	2,31
23	7,88	5,66	4,76	4,26	3,94	3,71	3,54	3,41	3,30	3,21	3,07	2,93	2,78	2,70	2,62	2,54	2,45	2,35	2,26
24	7,82	5,61	4,72	4,22	3,90	3,67	3,59	3,36	3,26	3,17	3,03	2,89	2,74	2,66	2,58	2,49	2,40	2,31	2,21
25	7,77	5,57	4,68	4,18	3,86	3,63	3,46	3,32	3,22	3,13	2,99	2,85	2,70	2,62	2,54	2,45	2,36	2,27	2,17
26	7,72	5,53	4,64	4,14	3,82	3,59	3,42	3,29	3,18	3,09	2,96	2,82	2,66	2,58	2,50	2,42	2,33	2,23	2,13
27	7,68	5,49	4,60	4,11	3,78	3,56	3,39	3,26	3,15	3,06	2,93	2,78	2,63	2,55	2,47	2,38	2,29	2,20	2,10
28	7,64	5,45	4,57	4,07	3,75	3,53	3,36	3,23	3,12	3,03	2,90	2,75	2,60	2,52	2,44	2,35	2,26	2,17	2,06
29	7,60	5,42	4,54	4,04	3,73	3,50	3,36	2,20	3,09	3,00	2,87	2,73	2,57	2,49	2,41	2,33	2,23	2,14	2,03
30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,07	2,98	2,84	2,70	2,55	2,47	2,39	2,30	2,21	2,11	2,01
40	7,31	5,18	4,81	3,83	3,51	3,29	3,12	2,99	2,89	2,80	2,66	2,52	2,37	2,99	2,99	2,11	2,02	1,92	1,80
60	7,08	4,98	4,13	3,65	3,34	3,12	2,95	2,82	2,72	2,63	2,50	2,35	2,20	2,12	2,03	1,94	1,84	1,73	1,60
120	6,85	4,79	3,95	3,48	3,17	2,96	2,79	2,66	2,56	2,47	2,34	2,19	2,03	1,95	1,86	1,76	1,66	1,53	1,38
∞	6,63	4,61	3,78	3,32	3,03	2,80	2,64	2,51	2,41	2,32	2,18	2,04	1,88	1,79	1,70	1,59	1,47	1,32	1,00

Source: E.S. Pearson and H.O. Hartley, *Biometrika Tables for Statisticians*, Vol.2 (1972), table 5, pages 178. Reprodution autoride

Bảng 5. BẢNG PHÂN BỐ COCHRAN

k \ m	Mức ý nghĩa α = 0,05													
	1	2	3	4	5	6	7	8	9	10	16	36	44	∞
2	0,99	9750	9392	9057	8872	8534	8332	8159	8010	7800	7341	6602	5813	5000
3	966	8790	7977	7557	7071	6771	6530	6333	6167	6025	5466	4748	4031	3333
4	906	7679	6841	6287	5895	5598	5365	5175	5017	4884	4366	2720	3093	2500
5	841	6838	5981	5441	5065	4783	4564	4387	4241	4118	3645	3066	2513	2000
6	780	6161	5321	4803	4447	4184	3980	3817	3682	3568	3135	2612	2119	1667
7	727	5612	4800	4307	3974	3726	3535	3384	3259	3154	2756	2278	1833	1429
8	679	5157	4377	3910	3595	3362	3185	3043	2926	2829	2462	2022	1616	1250
9	638	4775	4027	3584	3286	3067	2901	2768	2659	2568	2226	1820	1446	1111
10	602	4450	3733	3311	3029	2823	2666	2541	2439	2353	2032	2655	1380	1000
12	541	3924	3264	2880	2624	2493	2299	2187	2098	2020	1737	1403	1100	0833
15	470	3346	2750	2491	2195	2034	1911	1815	1736	1671	1429	1144	0889	0667
20	289	2705	2205	1921	1735	1602	1501	1422	1357	1303	1108	0879	0675	0500
24	243	2354	1907	1656	1493	1374	1286	1216	1160	1113	0942	1743	0567	0417
30	292	1980	1593	1377	1237	1137	1061	1002	0958	0921	0771	0604	0457	0333
40	237	1576	1259	1082	0968	0887	0827	0780	0745	0713	0595	0426	0374	0250
60	173	1131	0895	0765	0682	0623	0583	0552	0520	0497	0411	0316	0234	0167
120	099	0632	0945	0419	0371	0337	0312	0292	0279	0266	0218	0165	0120	0083
∞	000	0000	0000	0000	0000	0000	0000	0000	0000	0000	0000	0000	0000	0000

Bảng 6. BẢNG COCHRAN

k \ m	Mức ý nghĩa α = 0,01													
	1	2	3	4	5	6	7	8	9	10	16	36	144	∞
2	9999	9950	9794	9586	9373	9172	8998	8823	8674	8539	7949	7067	6062	5000
3	9933	9423	8831	8335	7933	7606	7335	7107	6912	6743	6059	5153	4230	3333
4	9676	8643	7814	7212	6761	6410	6129	5879	5702	5536	4884	0457	3251	2500
5	9279	7885	6957	6329	5875	5531	5259	5037	4854	4697	4094	3351	2644	2000
6	8828	7218	6258	5635	5195	4866	4608	4401	4229	4084	3529	2858	2229	1667
7	8376	6644	5685	5080	4659	4347	4105	3911	3751	3616	3105	2494	1929	1429
8	7945	6152	5209	4627	4226	3932	3704	3522	3373	3248	2779	2214	1700	1250
9	7544	5727	4810	4251	3870	3592	3378	3207	3067	2950	2514	1992	1521	1111
10	7175	5358	4469	3934	3572	3308	3106	2945	2813	2704	2297	1811	1376	1000
12	6528	4751	3919	3428	3099	2861	2680	2535	2419	2320	1961	1535	1157	0833
15	5747	4069	3317	2882	2593	2386	2228	2104	2002	1918	1612	1251	0934	0667
20	4799	3297	2654	2288	2048	1877	1748	1646	1567	1501	1248	0960	0709	0500
24	4247	2871	2295	1970	1759	1608	1495	1406	1338	1283	1060	0810	0595	0417
30	3632	2412	1913	1635	1454	1327	1232	1157	1100	1054	0864	0658	0480	0333
40	2940	1915	1508	1281	1135	1033	0957	0898	0853	0816	0668	0503	0363	0250
60	2151	1371	1069	0902	0796	0722	0668	0625	0594	0567	0461	0344	0245	0167
120	1225	0759	0585	0489	0429	0387	0357	0334	0316	0302	0242	0178	0125	0083
∞	0000	0000	0000	0000	0000	0000	0000	0000	0000	0000	0000	0000	0000	0000

Bảng 7. BẢNG "KHI BÌNH PHƯƠNG"

ϖ	$\chi^2_{,005}$	$\chi^2_{,01}$	$\chi^2_{,025}$	$\chi^2_{,05}$	$\chi^2_{,10}$	$\chi^2_{,25}$	$\chi^2_{,50}$	$\chi^2_{,75}$	$\chi^2_{,90}$	$\chi^2_{,95}$	$\chi^2_{,975}$	$\chi^2_{,99}$	$\chi^2_{,995}$	$\chi^2_{,999}$
1	,0000	,0010	,0039	,0158	,102	,455	1,32	2,71	3,84	5,02	6,63	7,83	10,8	
2	,0100	,0201	,0506	,103	,211	,575	1,39	2,77	4,61	5,99	7,38	9,21	10,6	13,8
3	,0717	,115	,216	,352	,584	1,21	2,37	4,11	6,25	7,81	9,35	11,3	12,8	16,3
4	,207	,297	,484	,711	1,06	1,92	3,36	5,39	7,78	9,49	11,1	13,3	14,9	18,5
5	,412	,554	,831	1,15	1,61	2,67	4,35	6,63	9,24	11,1	12,8	15,1	16,7	20,5
6	,676	,872	1,24	1,64	2,20	3,45	5,35	7,84	10,6	12,6	14,4	16,8	18,5	22,5
7	,989	1,24	1,69	2,17	2,83	4,25	6,35	9,04	12,0	14,1	16,0	18,5	20,3	24,3
8	1,34	1,65	2,18	2,73	3,49	5,07	7,34	10,2	13,4	15,5	17,5	20,1	22,0	26,1
9	1,73	2,09	2,70	3,33	4,17	5,90	8,34	11,4	14,7	16,9	19,0	21,7	23,6	27,9
10	2,16	2,56	2,25	3,94	4,87	6,74	9,34	12,5	16,0	18,3	20,5	23,2	25,2	29,6
11	2,60	3,05	3,82	4,57	5,58	7,58	10,3	13,7	17,3	19,7	24,9	24,7	26,8	31,3
12	3,07	3,57	3,40	5,23	6,30	8,44	11,3	14,8	18,1	21,0	23,3	26,2	28,3	32,9
13	3,57	4,11	5,01	5,89	7,04	9,30	12,3	16,0	19,8	22,4	24,7	27,7	29,8	34,5
14	4,07	4,66	5,63	6,57	7,79	10,2	13,3	17,1	21,1	23,7	26,1	29,1	31,3	36,1
15	4,60	5,23	6,26	7,26	8,55	11,0	14,3	18,2	22,3	25,0	27,5	30,6	32,8	37,7
16	5,14	5,81	6,91	7,96	9,31	11,9	15,3	19,4	23,5	26,3	28,8	32,0	34,3	39,3
17	5,70	6,41	7,56	8,67	10,1	12,8	16,3	20,5	24,8	27,6	30,2	33,4	35,7	46,8
18	6,26	7,01	8,23	9,39	10,9	13,7	17,3	21,6	26,0	28,9	31,5	34,8	37,2	42,3
19	6,84	7,63	8,91	10,1	11,7	14,6	18,3	22,7	27,2	30,1	32,9	36,2	38,6	43,8
20	7,43	8,26	9,59	10,9	12,4	15,5	19,3	23,8	28,4	31,4	34,2	37,6	40,0	45,3

Bảng 7 (tiếp theo)

ϖ	$\chi^2_{,005}$	$\chi^2_{,01}$	$\chi^2_{,025}$	$\chi^2_{,05}$	$\chi^2_{,10}$	$\chi^2_{,25}$	$\chi^2_{,50}$	$\chi^2_{,75}$	$\chi^2_{,90}$	$\chi^2_{,95}$	$\chi^2_{,975}$	$\chi^2_{,99}$	$\chi^2_{,995}$	$\chi^2_{,999}$
21	8,03	8,90	10,3	11,6	13,2	16,3	20,3	24,9	29,6	32,7	35,5	38,9	41,4	46,8
22	8,64	9,54	11,0	12,3	14,0	17,2	21,3	26,0	30,8	33,9	36,8	40,3	42,8	48,3
23	9,26	10,2	11,7	13,1	14,8	18,1	22,3	27,1	32,0	35,2	38,1	41,6	44,2	49,7
24	8,89	10,9	12,4	13,8	15,7	19,0	13,3	28,2	33,2	36,4	39,4	43,0	45,6	51,2
25	10,5	11,5	13,1	14,6	16,5	19,9	24,3	29,3	34,4	37,7	40,6	44,3	46,9	52,6
26	11,2	12,2	13,8	15,4	17,3	20,8	25,3	30,4	35,6	33,9	41,9	45,6	48,3	54,1
27	11,8	12,9	14,6	16,2	18,1	21,7	26,0	31,5	36,7	40,1	43,2	47,0	49,6	55,5
28	12,5	13,6	15,3	16,9	18,9	22,7	27,3	32,6	37,9	41,3	44,5	48,3	51,0	56,9
29	13,1	14,3	16,0	17,7	19,8	23,6	28,3	33,7	39,1	42,6	45,7	49,6	52,3	58,3
30	13,8	15,0	16,8	18,5	20,6	24,5	29,3	34,8	40,3	43,8	47,0	50,9	53,7	59,7
40	20,7	22,2	24,4	26,5	29,1	33,7	39,3	45,6	51,8	55,8	59,3	63,7	66,8	73,4
50	28,0	29,7	32,4	34,8	37,7	42,9	49,3	56,3	63,2	67,5	71,4	76,2	79,5	86,7
60	35,5	37,5	40,5	43,2	46,5	52,3	59,3	67,0	74,4	79,1	83,3	88,4	92,0	99,6
70	43,3	45,4	48,8	51,7	55,3	61,7	69,3	77,6	85,5	90,5	95,0	100	104	112
80	51,2	53,5	57,2	60,4	64,0	7101	79,3	88,1	96,6	102	107	112	116	125
90	59,2	61,8	65,6	69,1	73,3	80,6	89,3	98,6	108	113	118	124	128	137
100	67,3	70,1	74,2	77,9	82,4	90,1	99,3	109	118	124	130	136	140	149

Source: E.S. Pearson and H.O. Hartley, *Biometrika Tables for Statisticians*, Vol.1 (1966), table 8, pages 137 et 138. Reproduction autorisée.

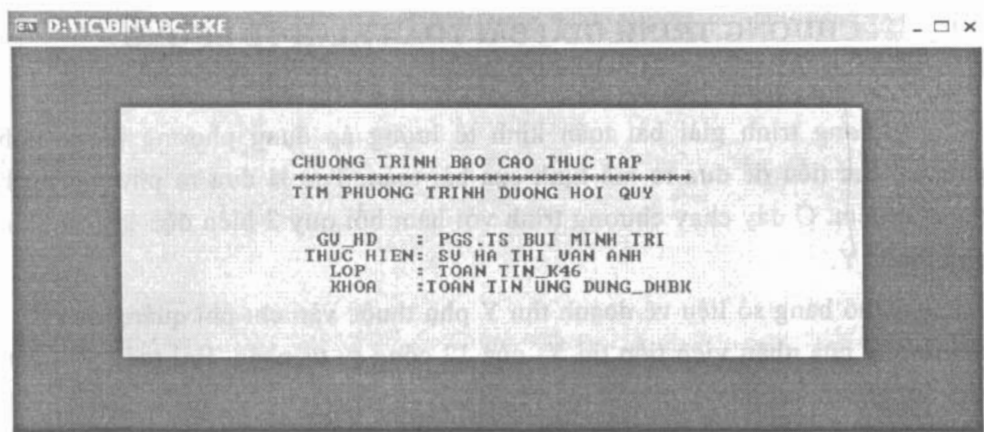
PHỤ LỤC 2
CHƯƠNG TRÌNH GIẢI BÀI TOÁN KINH TẾ LƯỢNG

Chương trình giải bài toán kinh tế lượng áp dụng phương pháp bình phương cực tiểu để đưa ra mô hình của bài toán. Tức là đưa ra phương trình thực nghiệm. Ở đây chạy chương trình với hàm hồi quy 2 biến độc lập và biến phụ thuộc Y.

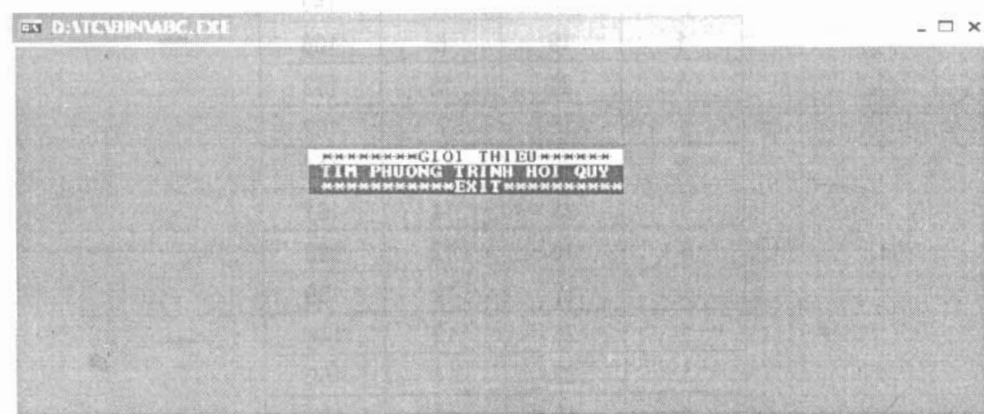
Cho bảng số liệu về doanh thu Y phụ thuộc vào chi phí quảng cáo X_1 , tiền lương của nhân viên tiếp thị X_2 của 12 công ty tư nhân. Bài toán yêu cầu xây dựng hàm hồi quy Y phụ thuộc vào X_1, X_2

STT	x_1	x_2	y
1	18	10	127
1	25	11	149
1	19	6	106
1	24	16	163
1	15	7	102
1	26	17	180
1	25	14	161
1	16	12	128
1	17	12	139
1	23	12	144
1	22	14	159
1	15	15	138

Giới thiệu chương trình:



Giao diện chính của chương trình:



Màn hình nhập số liệu:

Nhập ma trận X:

```
D:\TC\BIN\ABC.EXE
              X[4][1]=1     X[4][2]=24     X[4][3]=16
              X[5][1]=1     X[5][2]=15     X[5][3]=7
              X[6][1]=1     X[6][2]=26     X[6][3]=17
              X[7][1]=1     X[7][2]=25     X[7][3]=14
              X[8][1]=1     X[8][2]=16     X[8][3]=12
              X[9][1]=1     X[9][2]=17     X[9][3]=12
              X[10][1]=1    X[10][2]=23    X[10][3]=12
              X[11][1]=1    X[11][2]=22    X[11][3]=14
              X[12][1]=1    X[12][2]=15    X[12][3]=15

ma tran X la:
   1>            1.0000          18.0000          10.0000
   2>            1.0000          25.0000          11.0000
   3>            1.0000          19.0000           6.0000
   4>            1.0000          24.0000          16.0000
   5>            1.0000          15.0000           7.0000
   6>            1.0000          26.0000          17.0000
   7>            1.0000          25.0000          14.0000
   8>            1.0000          16.0000          12.0000
   9>            1.0000          17.0000          12.0000
  10>            1.0000          23.0000          12.0000
  11>            1.0000          22.0000          14.0000
  12>            1.0000          15.0000          15.0000

co can sua khong? c/k_
```

Nhập vector Y:

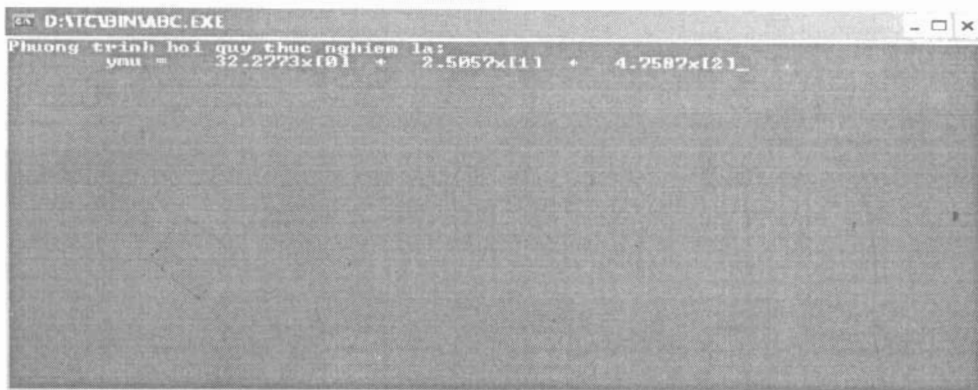
```
D:\TC\BIN\ABC.EXE
         Y[4]=163
         Y[5]=102
         Y[6]=180
         Y[7]=161
         Y[8]=128
         Y[9]=139
         Y[10]=144
         Y[11]=159
         Y[12]=138
vecto  Y la:
   1>          127.0000
   2>          149.0000
   3>          106.0000
   4>          163.0000
   5>          102.0000
   6>          180.0000
   7>          161.0000
   8>          128.0000
   9>          139.0000
  10>          144.0000
  11>          159.0000
  12>          138.0000

co can sua khong? c/k
```

Kết quả chương trình:

Phương trình thực nghiệm tìm ra là:

$$\tilde{y} = 32{,}2777 + 2{,}5057x_1 + 4{,}7587x_2$$



Xét với tệp có 4 biến quan sát và 54 mẫu: Tệp đó được lưu dưới tên file *mt54.txt*

1	0,29	5,31	0,94
1	–0,11	5,6	6,0
1	0,31	5,49	6,08
1	–0,19	5,8	6,17
1	–0,33	5,61	6,14
1	–0,09	5,28	6,09
1	–0,01	5,19	5,87
1	0,12	5,18	5,84
1	–0,07	5,3	5,99
1	0,41	5,23	6,12
1	–0,02	5,64	6,42
1	0,05	5,62	6,48
1	0,16	5,67	6,52
1	–0,35	5,83	6,64
1	0,23	5,53	6,75
1	0,33	5,76	6,73
1	0,43	6,09	6,89
1	0,16	6,52	6,98
1	0,39	6,68	6,98
1	0,05	7,07	7,1

1	0,13	7,12	7,19
1	0,60	7,25	7,29
1	0,17	7,85	7,65
1	–0,15	8,02	7,75
1	–0,73	7,87	7,72
1	0,06	7,14	7,67
1	0,39	7,20	7,66
1	0,15	7,59	7,89
1	–0,23	7,74	8,14
1	–0,0	7,51	8,21
1	–0,37	7,46	8,05
1	–0,27	7,09	7,94
1	–0,6	6,82	7,88
1	–0,61	6,22	7,79
1	–0,13	5,61	7,41
1	–0,7	5,48	7,18
1	–0,64	4,78	7,15
1	0,5	4,14	7,27
1	0,88	4,64	7,37
1	0,43	5,52	7,54
1	0,25	5,95	7,58
1	–0,17	6,20	7,62
1	–0,43	6,03	7,58
1	–0,34	5,60	7,48
1	–0,3	5,26	7,35
1	0,32	4,96	7,19
1	0,09	5,28	7,19
1	0,16	5,37	7,11
1	0,19	5,53	7,16
1	0,32	5,72	7,22
1	–0,38	6,04	7,36
1	0,09	5,66	7,34
1	0,07	5,75	7,3
1	0,08	5,82	7,3

Xét tệp chứa vector Y có 54 cột lưu trong file *vt54.txt*

1,146

−2,443

1,497

−0,132

2,025

0,737

−1,023

−0,956

0,385

0,983

5,092

3,649

2,703

−0,271

2,055

−0,714

0,653

−0,034

−1,058

−2,051

1,451

−0,989

1,358

0,746

1,855

−1,894

0,781

−1,161

2,233
2,425
2,169
0,982
4,708
6,063
9,382
9,304
10,690
6,531
7,873
3,882
4,960
1,301
1,154
0,116
4,928
2,530
8,425
5,291
5,192
0,257
4,402
3,173
5,104
4,646

Ma trận X là

```
D:\TC\BIN\ABC.EXE
vao ten tep chua mt can doc:mt54.txt
m=54    n=4
 ma tran X la:
    1>          1.0000          0.2900          5.3100          5.9400
    2>          1.0000         -0.1100          5.6000          6.0000
    3>          1.0000          0.3100          5.4900          6.0800
    4>          1.0000         -0.1900          5.8000          6.1700
    5>          1.0000         -0.3300          5.6100          6.1400
    6>          1.0000         -0.0900          5.2800          6.0900
    7>          1.0000         -0.0100          5.1900          5.8700
    8>          1.0000          0.1200          5.1800          5.8400
    9>          1.0000         -0.0700          5.3000          5.9900
   10>          1.0000          0.4100          5.2300          6.1200
   11>          1.0000         -0.0200          5.6400          6.4200
   12>          1.0000          0.0500          5.6200          6.4800
   13>          1.0000          0.1600          5.6700          6.5200
   14>          1.0000         -0.3000          5.8300          6.6400
   15>          1.0000          0.2300          5.5300          6.7500
   16>          1.0000          0.3300          5.7600          6.7300
   17>          1.0000          0.4300          6.0900          6.8900
   18>          1.0000          0.1600          6.5200          6.9800
   19>          1.0000          0.3900          6.6800          6.9800
   20>          1.0000          0.0500          7.0700          7.1000
nhan mot phim bat ky de xem tiep:
```

```
D:\TC\BIN\ABC.EXE
   19>          1.0000          0.3900          6.6800          6.9800
   20>          1.0000          0.0500          7.0700          7.1000
nhan mot phim bat ky de xem tiep:
   21>          1.0000          0.1300          7.1200          7.1900
   22>          1.0000          0.6000          7.2500          7.2900
   23>          1.0000          0.1700          7.8500          7.6500
   24>          1.0000         -0.1500          8.0200          7.7500
   25>          1.0000         -0.7300          7.8700          7.7200
   26>          1.0000          0.0600          7.1400          7.6700
   27>          1.0000          0.3900          7.2000          7.6600
   28>          1.0000          0.1500          7.5900          7.8900
   29>          1.0000         -0.2300          7.7400          8.1400
   30>          1.0000         -0.0500          7.5100          8.2100
   31>          1.0000         -0.3700          7.4600          8.0500
   32>          1.0000         -0.2700          7.0900          7.9400
   33>          1.0000         -0.6000          6.8200          7.8800
   34>          1.0000         -0.6100          6.2200          7.7900
   35>          1.0000         -0.1300          5.6100          7.4100
   36>          1.0000         -0.7000          5.4800          7.1800
   37>          1.0000         -0.6400          4.7800          7.1500
   38>          1.0000          0.5000          4.1400          7.2700
   39>          1.0000          0.8800          4.6400          7.3700
   40>          1.0000          0.4300          5.5200          7.5400
nhan mot phim bat ky de xem tiep:
```

```
D:\TC\BIN\ABC.EXE
   33>          1.0000         -0.6000          6.8200          7.8800
   34>          1.0000         -0.6100          6.2200          7.7900
   35>          1.0000         -0.1300          5.6100          7.4100
   36>          1.0000         -0.7000          5.4800          7.1800
   37>          1.0000         -0.6400          4.7800          7.1500
   38>          1.0000          0.5000          4.1400          7.2700
   39>          1.0000          0.8800          4.6400          7.3700
   40>          1.0000          0.4300          5.5200          7.5400
nhan mot phim bat ky de xem tiep:
   41>          1.0000          0.2500          5.9500          7.5800
   42>          1.0000         -0.1700          6.2000          7.6200
   43>          1.0000         -0.4300          6.0300          7.5800
   44>          1.0000         -0.3400          5.6000          7.4800
   45>          1.0000         -0.3000          5.2600          7.3500
   46>          1.0000          0.3200          4.9600          7.1900
   47>          1.0000          0.0900          5.2800          7.1900
   48>          1.0000          0.1600          5.3700          7.1100
   49>          1.0000          0.1900          5.5300          7.1600
   50>          1.0000          0.3200          5.7200          7.2200
   51>          1.0000         -0.3800          6.0400          7.3600
   52>          1.0000          0.0900          5.6600          7.3400
   53>          1.0000          0.0700          5.7500          7.3000
   54>          1.0000          0.0800          5.8200          7.3000

co can sua khong? c/k_
```

Vector Y:

```
D:\TC\BIN\ABC.EXE
vao ten tep chua vt can doc:vt54.txt
m=54   vecto  Y la:
   1>            1.1460
   2>           -2.4430
   3>            1.4970
   4>           -0.1320
   5>            2.0250
   6>            0.7370
   7>           -1.0230
   8>           -0.9560
   9>            0.3850
  10>            0.9830
  11>            5.0920
  12>            3.6490
  13>            2.7030
  14>           -0.2710
  15>            2.0550
  16>           -0.7140
  17>            0.6530
  18>           -0.0340
  19>           -1.0580
  20>           -2.0510
nhan phin bat ky de xem tiep...
```

```
D:\TC\BIN\ABC.EXE
  19>           -1.0580
  20>           -2.0510
nhan phin bat ky de xem tiep...
  21>            1.4510
  22>           -0.9890
  23>            1.3580
  24>            0.7460
  25>            1.8550
  26>           -1.8940
  27>            0.7810
  28>           -1.1610
  29>            2.2330
  30>            2.4250
  31>            2.1690
  32>            0.9820
  33>            4.7080
  34>            6.0630
  35>            9.3820
  36>            9.3040
  37>           10.6900
  38>            6.5310
  39>            7.8730
  40>            3.8820
nhan phin bat ky de xem tiep...
```

```
D:\TC\BIN\ABC.EXE
  34>            6.0630
  35>            9.3820
  36>            9.3040
  37>           10.6900
  38>            6.5310
  39>            7.8730
  40>            3.8820
nhan phin bat ky de xem tiep...
  41>            4.9600
  42>            1.3010
  43>            1.1540
  44>            0.1160
  45>            4.9280
  46>            2.5300
  47>            8.4250
  48>            5.2910
  49>            5.1920
  50>            0.2570
  51>            4.4020
  52>            3.1730
  53>            5.1040
  54>            4.6460

co can sua khong? c/k
```

Cuối cùng ta tìm được phương trình hồi quy thực nghiệm:

$$\tilde{y} = -4,1218 - 1,9969x_1 - 2,8757x_2 + 3,3793x_3$$

CODE của chương trình Phụ Lục 2

```
#include<stdio.h>
#include <dos.h>
#include <ctype.h>
#include<conio.h>
#include<math.h>
#include<stdlib.h>

#define N 100           // So mau quan sat
#define K 100           // So bien doc lap

typedef float mt_nk[N][K]; // Ma tran thiet ke X
typedef float mt_kn[K][N]; // Ma tran chuyen vi cua ma
                              tran X: XT
typedef float mt_kk[K][K]; // Ma tran C
typedef float vt_n[1000];  // Bien phu thuoc Y
typedef float vt_k[K];     // Cac he so
vt_n y;
int n,k;

void Noi_dung();
void hide();//dau con tro
void show();//hien con tro
void cuaso(int maunen,int mauchu,int x1,int y1,int
x2,int y2);
int getkey();
void noidungmenu(char *s[]);
void menu(int x,int y,char *m[],int sh,int *chon);

void nhapkt_mt();
void nhap_mt(mt_nk X,char x);
void suamt(mt_nk X,char x);
void inmt(mt_nk X,char x);
void inmt(mt_kk C,char x);
void nhap_vt(vt_n y,char x);
```

```
        void invt(vt_n y);
        void suavt(vt_n y,char x);
        void doc_mt(int *n1,int *k1,mt_nk a1);
        void doc_vt(int *n1,vt_n b1);
        void docdulieu(mt_nk x,vt_n y);
        void Tim_he_so(mt_nk X,vt_n y);
        void giai_tgt(mt_nk X,vt_n y,vt_k x1);
        void He_so(mt_nk X,vt_k y);

 void main()
{

        mt_nk X;
        vt_n y;
        char *nk[10];int i,j,cho=0;
        clrscr();
        textmode(C80);
        textcolor(1);
        textbackground(6);
        window(10,5,70,20); clrscr();
        gotoxy(15,4);printf("CHUONG TRINH BAO CAO THUC
TAP ");
        gotoxy(15,5);printf("***********************************");
        gotoxy(15,6);printf("TIM PHUONG TRINH DUONG HOI
QUY");
        gotoxy(15,9);printf("GV_HD: PGS.TS BUI MINH TRI  ");
        gotoxy(15,10);printf("THUC HIEN: SV HA THI VAN
ANH");
        gotoxy(15,11);printf("   LOP  : TOAN TIN_K46");
        gotoxy(15,12);printf("   KHOA  :TOAN TIN UNG
DUNG_DHBK");
        hide();

        getch();clrscr();
   H: noidungmenu(nk);
```

```
	textmode(C80);
	textbackground(4);
	clrscr();
	menu(25,8,nk,3,&cho);
	clrscr();
	 switch(cho)
	{
		case 0:
			clrscr();
			Noi_dung();
			clrscr();
			goto H;
		case 1:
			clrscr();	// cuaso(1,15,1,1,80,25);
			He_so(X,y);
			clrscr();
			goto H;
		case 2:
			clrscr();
			return;
	}
	getch();
}

void hide()
{
   union REGS r;
   r.h.ah=1;
   r.x.cx=0x2000;
   int86(0x10,&r,&r); /*dau con tro*/
}
/*--------------------------------*/
void show()
{
	union REGS r;
```

```
        r.h.ah=1;
        r.x.cx=0x0607;
        int86(0x10,&r,&r);/*hien con tro*/
}
/*---------------------------------*/
void cuaso(int maunen,int mauchu,int x1,int y1,int x2,int y2)
{
  textbackground(maunen);
  textcolor(mauchu);
  window(x1,y1,x2,y2);
}
/*----------------------------------*/
 int getkey()
 {
 int key ;
 if ((key=getch())==0)
 key=getch();
 return key;
 }
/*-------------------------------------------*/
void menu(int x,int y,char *m[],int sh,int *chon)
 {
 int k,i;
 hide(); /*dau con tro*/
 for (i=0;i<sh;i++)
 {
  if (i==*chon)
  {
  gotoxy(x,y+i);
  textattr(0x71); /*74 mau chu*/
  cputs(m[i]);
  }
  else
        {
        gotoxy(x,y+i);
```

```
      textattr(0x1f);/*0x1f hop chay */
      cputs(m[i]);
      }
}
i=*chon;
while(1)
 {
 k=getkey();
switch(k)
 {
 case 72:
       gotoxy(x,y+i);
       textattr(0x1f);/*0x1f hop chay */
       cputs(m[i]);
       if(i==0)
       {
       i=sh-1;
       }
       else
       i--;
       gotoxy(x,y+i);
       textattr(0X71);/*74 mau chu*/
       cputs(m[i]);
       break;
 case 80:
              gotoxy(x,i+y);
              textattr(0x1f);/*0x1f hop chay */
              cputs(m[i]);
              if(i==sh-1)
              {
              i=0;
              }
              else
              i++;
              gotoxy(x,y+i);
```

```
        textattr(0x71); /*74 mau chu */
        cputs(m[i]);
        break;
  case 13:
        show();/*hien con tro*/
        cuaso(2,15,1,1,80,25);
        *chon=i;
        return;
  case 27:
        show(); /*hien con tro*/
        *chon=-1;
        return;
  }
  }
}
/***************************************/
void noidungmenu(char *s[])
{
 s[0]=" ********GIOI THIEU****** ";
 s[1]=" TIM PHUONG TRINH HOI QUY ";
 s[2]=" **********EXIT********** ";
}


 /**************************************/
void Tim_he_so(mt_nk X,vt_n y)
{
 FILE *f;
 int i,j,p;
 float t,alfa,beta,lamda,gama,muy;
 vt_n v;
 fflush(stdin);
 f=fopen("VD.txt","wt");
 for(p=1;p<=k;p++)  //bien doi ma tran a
        {
```

```
fprintf(f, "\nLap lan %d\n", k);
t=0;
muy=0;

for(i=p;i<=n;i++)
{
        v[i]=X[i][p];
        t+=v[i]*v[i];
}
alfa=sqrt(t);
fprintf(f, "alfa=%f",alfa);
fprintf(f, "\n");
lamda=X[p][p];
if(lamda>0)
        alfa=-alfa;
beta=alfa*alfa-alfa*lamda;
fprintf(f, "beta=%f",beta);
fprintf(f, "\n");
v[p]=v[p]-alfa;
for(j=p;j<=k;j++)
{
                muy=0;
                for(i=p;i<=n;i++)
                        muy+=v[i]*X[i][j];
                gama=muy/beta;
                fprintf(f, "gama=%f",gama);
                fprintf(f, "\n");
                for(i=p;i<=n;i++)
                        X[i][j]=X[i][j]-v[i]*gama;
}
getch();

muy=0;
for(i=p;i<=n;i++)
        muy+=v[i]*y[i];
```

```
        gama=muy/beta;
        for(i=p;i<=n;i++)
                y[i]=y[i]-v[i]*gama;
        // getch();
        for(i=1;i<=n;i++)
        {
                fprintf(f, "%4d>  ",i);
                for(j=1;j<=k;j++)
                {
                        if(fabs(X[i][j])==0)
                                X[i][j]=fabs(X[i][j]);
                        fprintf(f, "%14.5f",X[i][j]);
                }
                fprintf(f, "\n");
        }
         for(i=1;i<=n;i++)
         {
                fprintf(f, "%4d>  ",i);
                fprintf(f, "%14.5f",y[i]);
                fprintf(f, "\n");
         }
        }
        for (i=k+1;i<=n;i++)
        t+=y[i]*y[i];
        fprintf(f, "\n binh phuong do khong khop cua nghiem
la:%14.4f\n",t);
  fclose(f);
}

/*************************************************/
void nhapkt_mt()
        {
        nhap:
                printf("\n so mau quan sat n = ");
                scanf("%d",&n);
```

```
		printf("\nSo bien doc lap k = ");
		scanf("%d",&k);
		if (n<k)
		{
			printf("Bai toan khong co nghiem xac dinh;
moi nhap lai!");
			goto nhap ;
		}
	}
/**************************************************/
void nhap_mt(mt_nk X,char x)
	{
	int i,j;
	float tg;
	clrscr();
	printf("\n nhap ma tran X:\n");
	for (i=1;i<=n;i++)
	 {
		for (j=1;j<=k;j++)
		 {
			gotoxy(1+13*j,i+3);
			printf("%c[%d][%d]= ",x,i,j);
			scanf("%f",&tg);
			X[i][j]=tg;
		 }

	 }
	}
/*******************************************/
void nhap_vt(vt_n y,char x)
	{
	int i;
		printf("\n Moi nhap vecto %c:\n",x);
		for (i=1;i<=n;i++)
		 {
```

```
                    printf("\t %c[%d]=",x,i);
                    scanf("%f",&y[i]);
              }
       }
/*****************************************/
void inmt(mt_nk X,char x)
       {
       int i,j;
             printf("\n ma tran %c la: \n ",x);
             for (i=1;i<=n;i++)
              {  printf("%4d>",i);
                    for (j=1;j<=k;j++)
                    {
                          if (fabs(X[i][j])==0)
                          X[i][j]=abs(X[i][j]);
                          printf("\t%14.4f",X[i][j]);
                     }

                     printf("\n");
                     if(i%20==0)
                    {
                          printf("nhan mot phim bat ky de xem
tiep:\n");
                          getch();
                     }
               }
       }
/*****************************************/
void suamt(mt_nk X,char x)
   {
       int i,j;
       float tg;
       char tl;
        do
         {
```

```
            inmt(X,'X');
            printf ("\nco can sua khong? c/k");
            tl=getch();
            if (tl=='c'||tl=='C')
              {
                    printf("\nhang can sua:") ;
                    scanf("%d",&i);
                    printf("\ncot can sua:") ;
                    scanf("%d",&j);
                    printf("phan tu can sua
%c[%d][%d]=",x,i,j);
                    scanf("%f",&tg);
                    X[i][j]=tg;
                }
          }
        while (tl!='k');
      }
/****************************************/
void invt(vt_n y)
  {
        int i;

            for (i=1;i<=n;i++)
              {   printf("%4d>",i);
                    printf("\t%14.4f",y[i]);
                    printf("\n");
                    if (i%20==0)
                    {
                          printf("nhan phin bat ky de xem
tiep...\n");
                          getch();
                    }
              }
      }
/**************************************/
```

```
void suavt(vt_n y,char x)
  {
        int i;
        char tl;
        do
        {
          printf("vecto %c la:\n",x);
          invt(y);
          printf ("\nco can sua khong? c/k\n");
          tl=getch();
          if (tl=='c'||tl=='C')
           {
                printf("\n vi tri can sua:");
                scanf("%d",&i);
                printf("phan tu can sua %c[%d]=",x,i);
                scanf("\t%f",&y[i]);
                printf("\n");
             }
         }
         while (tl!='k');
}
/*****************************************/
void doc_mt(int *n,int *k,mt_nk X)
{
        int i,j;
        FILE *f1;
        char u,tep[15];
        clrscr();
        lai:
        printf("\nvao ten tep chua mt can doc:");
        gets(tep);
        fflush(stdin);
        f1=fopen(tep, "rt");
        if(f1==NULL)
        {
```

```
            printf("\nloi mo tep!");
            printf("\nnhan phim bat ky de nhap lai... ");
            printf("\nnhan 'e' de thoat... ");
            if(getch()=='e')
                  exit(1);
            else
            goto lai;
      }
      else
      {
      fscanf(f1, "%d%d%2c",n,k,u);
      for(i=1;i<=*n;i++)
      {
            for(j=1;j<=*k;j++)
            {
                  fscanf(f1, "%f",&X[i][j]);
            }
            fscanf(f1, "%2c",u);
      }
      }
      fclose(f1);
      getch();
}
/***************************************/
void doc_vt(int *n,vt_n y)
{
      int i;
      FILE *f2;
      char u,tep[15];
      nhap_lai:
      clrscr();
      printf("vao ten tep chua vt can doc:");
      gets(tep);
      fflush(stdin);
      f2=fopen(tep, "rt");
```

```
	if(f2==NULL)
	{
	  printf("loi mo tep!!!");
	  printf("nhan 'e' de thoat_ nhan phim bat ky de nhap
lai... ");
	  if(getch()=='e')
		exit(1);
	  else
		goto nhap_lai;
	}
	else
	{
	fscanf(f2, "%d%2c",n,u);
	for(i=1;i<=*n;i++)
		fscanf(f2, "%f%2c",&y[i],u);
	}
	fclose(f2);
}
/*********************************************/
void docdulieu(mt_nk X,vt_n y)
{
	char cha,chon;
	clrscr();

	printf("\nban muon nhap dl tu tep khong?(c/k): ");
	chon=getch();
	putch(chon);
	getch();
	if(chon!='c')
	 {
		nhapkt_mt();
		nhap_mt(X, 'X');
		suamt(X, 'X');
		getch();
		clrscr();
```

```
		nhap_vt(y,'Y');
		suavt(y,'Y');
		getch();
		clrscr();
	}
	else
	{
		doc_mt(&n,&k,X);
		printf("n=%d   ",n);
		printf("k=%d",k);
		getch();
		suamt(X,'X');
		doc_vt(&n,y);
		printf("n=%d   ",n);
		getch();
		suavt(y,'Y');
	}
}
/*******************************************/
void giai_tgt(mt_nk X,vt_n y,vt_k x)  //lkl
{
	FILE *f;
	int i,j;
	float t;
	f=fopen("KQ_househ.txt","at");
	fseek(f,0,2);
	fprintf(f,"Nghiem bai toan binh phuong cuc tieu:\n");
	for(i=k+1;i<=n;i++)
		x[i]=0;
	x[k]=y[k]/X[k][k];
	for(i=k-1;i>=1;i--)
	{
		t=0;
		for(j=k;j>=i+1;j--)
		{
```

```
                    t+=X[i][j]*x[j];
              }
              x[i]=(y[i]-t)/X[i][i];
       }
       clrscr();
       printf("Phuong trinh hoi quy thuc nghiem la:\n");
       printf("\tymu = ");
       for(i=1;i<=k-1;i++)
              {
                     if (fabs(x[i])==0)
                            x[i]=fabs(x[i]);
                     printf("\t(%8.4f)x[%d] + ",x[i],i-1);
              }
              printf("\t(%8.4f)x[%d]",x[k],k-1);
              //getch();
       for(i=1;i<=k;i++)
              {
                     fprintf(f,"%4d>  ",i);
                     fprintf(f,"%14.5f",x[i]);
                     fprintf(f,"\n");
              }

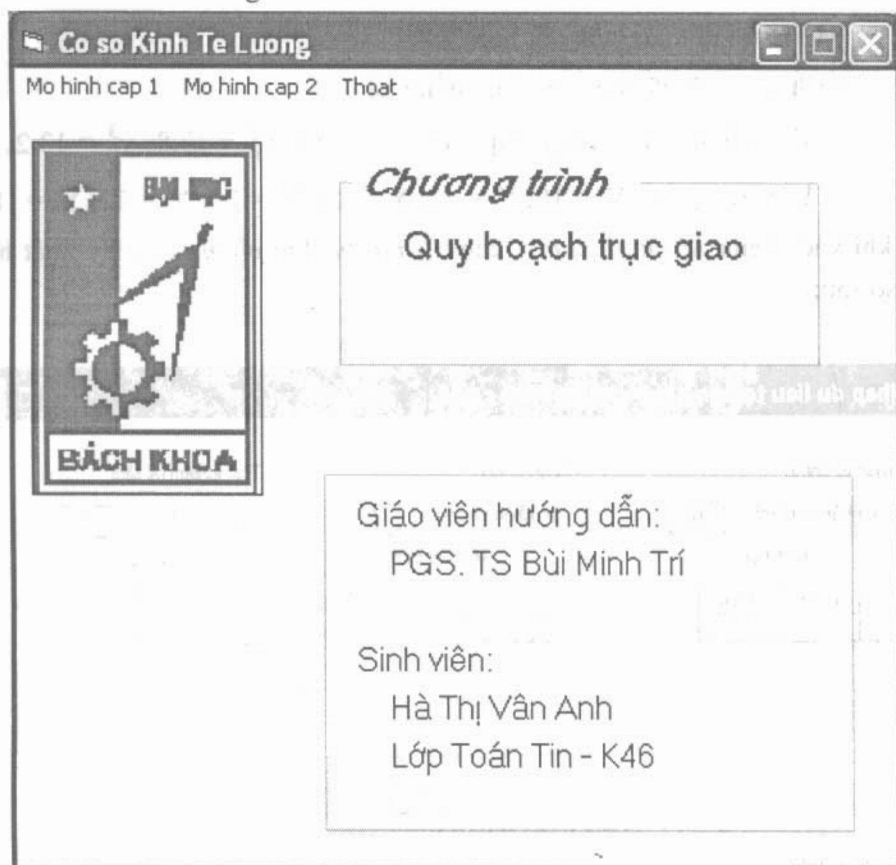
       fclose(f);
}
/**************************************************/
void He_so(mt_nk X,vt_n y )
  {
  char cha ;
  vt_k z; vt_n r;float gama;
  int i;
   do
   {
        docdulieu(X,y);
        Tim_he_so(X,y);
        giai_tgt(X,y,z);
```

```
        printf("\nban co muon nua khong?(c/k): ");
        cha=getch();
        }
        while(cha!='k');
   }

/********************************************/
void Noi_dung()
  {
        printf("Dau vao: \n Gia su x1,x2,...,xk la cac bien doc
lap va bien y la bien phu thuoc. ");
        printf("\nTa tien hanh n quan sat dong thoi ve k+1 bien
x1,x2,...xk, va y. ");
        printf("\nNoi dung: Bang phuong phap cuc tieu ta tim
ra cac he so phuong trinh thuc nghiem. ");
        printf("Da ra: \n Phuong trinh thuc nghiem");
        getch();
  }
```

PHỤ LỤC 3
BÁO CÁO CHƯƠNG TRÌNH

Menu chính của chương trình:



Sau khi chuyển con chuột đến thanh menu trên cùng. Lựa chọn loại mô hình quy hoạch trực giao.

Có ba Menu là:

- *Mo hinh cap 1*
- *Mo hinh cap 2*
- *Thoát*

1.1. Lựa chọn với mô hình cấp 1:

1.1.1. Nhập từ bàn phím

Xét ví dụ sau: dùng QHTG cấp 1 tìm quan hệ giữa Y và 3 biến Z_1, Z_2, Z_3

$$150 \le Z_1 \le 300$$

$$30 \le Z_2 \le 90$$

$$15 \le Z_3 \le 45$$

+ Giả sử mô hình là tuyến tính

+ Chọn $\alpha = 0,05$ cho các bảng thống kê

+) Tiến hành 3 thí nghiệm lặp ở tâm $y_0^1 = 12$; $y_0^2 = 13,8$; $y_0^3 = 13,2$.

+) $Y_1 = 3$; $Y_2 = 6$; $Y_3 = 10$; $Y_4 = 12$; $Y_5 = 15$; $Y_6 = 23$; $Y_7 = 12$; $Y_8 = 18$

Sau khi vào Menu *Mo hinh cap 1*, chọn *Nhap tu Ban phim*. Ta thấy xuất hiện cửa sổ sau:

Nhap du lieu tu ban phim cho mo hinh cap I

Nhân tố chính
Số nhân tố chính 0
Giới hạn
Dưới | Trên

Nhân tố phụ
Số nhân tố phụ 0
Hệ thức sinh

Thí nghiệm tại tâm
Số thí nghiệm 0
Kết quả đo

Build

Bảng thí nghiệm

Cancel Save OK

Trong thí dụ trên: Số nhân tố chính là 3, số nhân tố phụ là 0. Lần lượt nhập giới hạn trên, giới hạn dưới cho Z , nhập số thí nghiệm tại tâm và giá trị của nó.

Sau khi nhập xong ta Ấn vào nút ***Buid*** để nhập các giá trị cho Y.

Nhap du lieu tu ban phim cho mo hinh cap I

Nhân tố chính

Số nhân tố chính 3

Giới hạn

	Dưới	Trên
	150	300
	30	90
	15	90

Nhân tố phụ

Số nhân tố phụ 0

Hệ thức sinh

Thí nghiệm tại tâm

Số thí nghiệm 3

Kết quả đo

	12
	13.8
	13.2

Build

Bảng thí nghiệm

	x0	x1	x2	x3	y
2	+	+	-	-	6
3	+	-	+	-	10
4	+	+	+	-	12
5	+	-	-	+	15
6	+	+	-	+	23
7	+	-	+	+	18
8	+	+	+	+	18

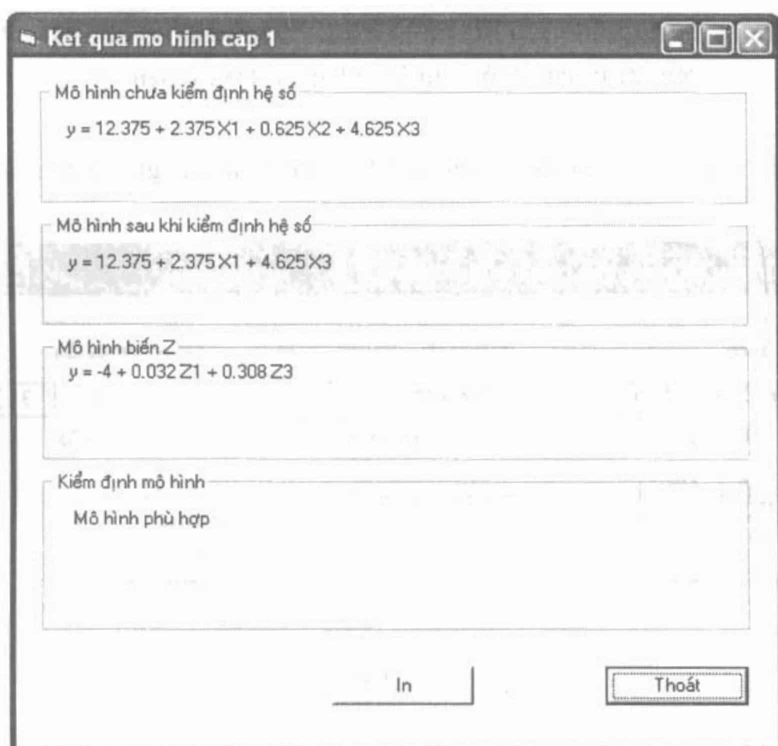
Cancel Save OK

+) Nếu muốn thực hiện chương trình Ấn ***Ok.***

+) Nếu không muốn thực hiện chương trình và quay trở lại thì Ấn Cancel.

+) Muốn ghi kết quả lại Ấn ***Save***

Sau khi Ấn ***Ok***, kết quả như sau:



1.1.2. Nhập từ File

Kích vào *Nhap tu File* sẽ hiện lên hộp thoại.



Chọn File cần lấy dữ liệu là Vi du 4.1. Rồi ấn ***Open*** trên màn hình sẽ như sau:

Nhap du lieu tu ban phim cho mo hinh cap I

Nhân tố chính

Số nhân tố chính: 3

Giới hạn

	Dưới	Trên
1	0.008	0.1
2	0.03	0.3
3	19	84

Nhân tố phụ

Số nhân tố phụ: 0

Hệ thức sinh

Thí nghiệm tại tâm

Số thí nghiệm: 3

Kết quả đo

1	11.1
2	11.3
3	10.8

Bảng thí nghiệm

	x0	x1	x2	x3	y
1	+	-	-	-	13.9
2	+	+	-	-	18.5
3	+	-	+	-	2
4	+	+	+	-	3
5	+	-	-	+	16
6	+	+	-	+	18.5
7	+	-	+	+	9

Cancel OK

+) Ấn ***Ok*** nếu muốn thực hiện tiếp

+) Ấn ***Cancel*** nếu không muốn thực hiện nữa.

Sau khi ấn ***Ok***, kết quả hiện lên như sau:

Ket qua mo hinh cap 1

Mô hình chưa kiểm định hệ số

y = 12.375 + 2.375X1 + 0.625X2 + 4.625X3

Mô hình sau khi kiểm định hệ số

y = 12.375 + 2.375X1 + 4.625X3

Mô hình biến Z

y = -4 + 0.032Z1 + 0.308Z3

Kiểm định mô hình

Mô hình phù hợp

In Thoát

2.2. Lựa chọn với mô hình cấp 2

2.2.1. Nhập từ bàn phím

Xét ví dụ sau: Tìm quan hệ giữa Y và 3 biến Z_1, Z_2, Z_3

$$0,008 \le Z_1 \le 0,1$$

$$0,03 \le Z_2 \le 0,3$$

$$19 \le Z_1 \le 84$$

+ Giả sử mô hình là tuyến tính

+ Chọn α=0,05 cho các bảng thống kê

+) Tiến hành 4 thí nghiệm lặp ở tâm

$y_0^1 = 10,1;\ y_0^2 = 11,2;\ y_0^3 = 9,9;\ y_0^4 = 12,3$

+) $Y_1 = 13,9$; $Y_2 = 18,5$; $Y_3 = 2$; $Y_4 = 3$; $Y_5 = 16$; $Y_6 = 18.5$; $Y_7 = 9$; $Y_8 = 12$

Nhap du lieu tu ban phim cho mo hinh cap II

Nhân tố chính

Số nhân tố chính 3

Giới hạn

Dưới	Trên
0.008	0.1
0.03	0.3
19	84

Thí nghiệm tại tâm lần 1

Số thí nghiệm 4

Thí nghiệm tại tâm lần 2

Số thí nghiệm 4

Kết quả đo

11.2
9.9
12.3

Build

Bảng thí nghiệm

	x1	x2	x3	x1 x2	x1 x3	x2 x3	x'1	x'2	x'3	y
14	-1.414	0	0	0	0	0	1.333	-0.667	-0.667	8
15	0	1.414	0	0	0	0	-0.667	1.333	-0.667	7.5
16	0	-1.414	0	0	0	0	-0.667	1.333	-0.667	15.8
17	0	0	1.414	0	0	0	-0.667	-0.667	1.333	11.5
18	0	0	-1.414	0	0	0	-0.667	-0.667	1.333	5

Cancel Save OK

+) Ấn ***Ok*** nếu muốn thực hiện tiếp.

+) Ấn ***Cancel*** nếu không muốn thực hiện nữa.

Sau khi ấn ***Ok***, kết quả hiện lên như sau:

Ket qua mo hinh cap 2

Mô hình chưa kiểm định hệ số

y = 11.067 + 1.75X1 - 4.386X2 + 2.274X3 - 0.387X1X2 - 0.013X1X3 + 1.737X2X3 + 0.762X'1^2 + 0.837X'2^2 - 0.863X'3^2

Mô hình sau khi kiểm định hệ số

y = 11.067 + 1.75X1 - 4.386X2 + 2.274X3 + 1.737X2X3

Mô hình biến Z

y = 14.135 + 38.043Z1 - 52.887Z2 + 0.005Z3 + 0.396Z2Z3

Kiểm định mô hình

Mô hình bậc hai phù hợp

In | Thoát

2.2.2. Nhập từ File tương tự như trong QHTG cấp 2

2.3. Thoát

Khi vào Menu này, tức là thoát chương trình .

TÀI LIỆU THAM KHẢO

[1] *Yu Adler.* Nhập môn Quy hoạch thực nghiệm (tiếng Nga). Moskva 1974.

[2] *L. J. Bain, M. Engelhard.* Introduction to Probability and Mathematical Startistics (Prindle Weber – Schmidt 1987).

[3] *W. Barnes* Statistical analysis for Engineer and Scientists; McGraw Hill International Editions, Printed in Singapor 1994.

[4] *W. G. Cochran - G. M. Cox* Experimental Designs, Znd ed, Wiley, NY 1957.

[5] *Bùi Công Cường - Bùi Minh Trí.* Xác suất thống kê ứng dụng; NXB Giao thông vận tải, Hà Nội 1997.

[6] *D. Dacunha - Castelle – M. Duflo* Probabilités et Statistiques (Paris 1994).

[7] *Trần Tuấn Điệp, Lý Hoàng Tú.* Giáo trình lý thuyết xác suất và thống kê toán học; NXB ĐH và THCN, Hà Nội 1979.

[8] *V. V. Federov.* Lý thuyết thực nghiệm tối ưu (tiếng Nga); Moskva 1971.

[9] *W. Feller An.* Introduction to Probability Theory and its Application; Znd ed Wiley NY 1971.

[10] *B. V. Gnedenko.* Giáo trình lý thuyết xác suất; NXB Toán Lý Moskva 1961 (Tiếng Nga).

[11] *C. R. Hicks* Fundamental Concepts in the Design of Experiments 3rd Ed; Rinehort and Winston NY 1982.

[12] *Đào Hữu Hồ, Nguyễn Văn Hữu, Hoàng Hữu Như.* Thống kê Toán học; NXB ĐH và THCN, Hà Nội 1984.

[13] *Nguyễn Văn Hộ.* Xác suất thống kê; NXB Giáo dục, Hà Nội 2001.

[14] *V. V. Kafarov.* Các phương pháp điều khiển trong công nghệ hoá học (tiếng Nga) Maskva 1971.

[15] *Bach Quoc Khang.* Texnologicheskie obosnovanie processov obmodota i separatxi risa aksialno- selariuiusimi utroistvami; Dissertatxi Doktora texn. Nayk. Krasnodar 1994.

[16] *Khoa Toán - Tin ứng dụng.* Bài giảng xác suất thống kê; Trường ĐHBK Hà Nội xuất bản 1998.

[17] *Koxtexki V. I.* Ma sát, bôi trơn và hao mòn trong máy móc; NXB Khoa học và Kỹ thuật Hà Nội 2003.

[18] *Phạm Văn Lang, Bạch Quốc Khang.* Cơ sở Lý thuyết Quy hoạch thực nghiệm và ứng dụng trong kỹ thuật nông nghiệp; NXB Nông nghiệp, Hà

Nội 1998.

[19] *R. J. Larsen, M. L. Marx Satistics*. Prentice Hall, New Jersey 1990.

[20] *M. Loeve*. Probability Theory 3rd ed; Van Nostrand Reinhold, NY 1963.

[21] *Nguyễn Văn Mạnh, Bùi Minh Trí*. Method of "cleft - overstep" by perpendicular direction for solving the unconstrained nonlinear optimization problem; Acta Mathematica Vietnamica, volume 15- 1990, N^0 2, pp 73-83.

[22] *W. Mendentall*. Introduction to Probability and Statistics 7th ed; Duxbury Press, Boston 1979.

[23] *B. N. Moixuk*. Các cơ sở của Lý thuyết thực nghiệm tối ưu (tiếng Nga), Moskva 1975.

[24] *B. N. Nalimov*. Lý thuyết các thực nghiệm (tiếng Nga); Moskva 1971.

[25] *Tống Đình Quỳ*. Giáo trình xác suất thống kê; NXB Giáo dục, Hà Nội 1999.

[26] *Bùi Minh Trí*. Mô hình toán kinh tế; NXB Khoa học và Kỹ thuật Hà Nội 2003.

[27] *B. J. Winer* Statistical Principles in Experimental Design 2nd ed; Mc GrawHill, NY 1971.

[28] *Bùi Minh Trí*. Điều khiển kinh tế; NXB Khoa học và Kỹ thuật Hà Nội, 2002.

[29] *Bùi Minh Trí*. Xác suất thống kê và quy hoạch thực nghiệm; NXB Khoa học và Kỹ thuật Hà Nội, 2004.

[30] *Nguyễn Quang Dong*. Bài giảng Kinh tế lượng, NXB Thống kê Hà Nội, 2005.

[31] *Nguyễn Văn Hữu, Nguyễn Hữu Dư*. Phân tích thống kê và dự báo; NXB Đại học Quốc gia Hà nội, 2003.

[32] *Nguyễn Hồ Quỳnh*. Chuỗi thời gian; Khoa Toán ứng dụng, Trường Đại học Bách khoa Hà Nội, 2001.

[33] *Montgomery, D. C., Peck, E. A. and Vining, G. G.* Introduction to Linear Regression; Analysis. 3rd Edition; New York, New York: John Wiley & Sons, 2001.

[34] *Arthur S. Goldberger*. Econometric Theory; John Wiley and Sons, Inc.

[35] *Damodar N. Gujarati*. Basic Econometrics, MacGraw-Hill, Inc, Thirt Ed 1995.

[36] *Madala, G. S. Macmillan*. Introduction of Econometrics New York 1992.

MỤC LỤC

Chương 3
HỒI QUY BỘI

Chương 4

MỘT SỐ TRƯỜNG HỢP MỞ RỘNG CỦA HỒI QUY BỘI

Chương 5

QUY HOẠCH TRỰC GIAO

Chương 6

ÁP DỤNG MÔ HÌNH HỒI QUY VÀO PHƯƠNG PHÁP NGOẠI SUY ĐỂ DỰ BÁO

PHỤ LỤC

TRƯỜNG ĐẠI HỌC BÁCH KHOA HÀ NỘI

KINH TẾ LƯỢNG

Tác giả: PGS. TS. Bùi Minh Trí

Chịu trách nhiệm xuất bản: PGS. TS. TÔ ĐĂNG HẢI
Biên tập và sửa bài: ThS. NGUYỄN HUY TIẾN
TRUNG KIÊN – QUANG NGỌC
Trình bày bìa: HƯƠNG LAN

NHÀ XUẤT BẢN KHOA HỌC VÀ KỸ THUẬT
70 Trần Hưng Đạo - Hà Nội

In 800 cuốn, khổ 16 × 24 cm, tại Xưởng in NXB Văn hoá Dân tộc
Quyết định xuất bản số: 539–2006/CXB/2–45/KHKT–10/8/2006
In xong và nộp lưu chiểu quý III năm 2006.

206298



Giá: 30.000đ